

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 534 180**

51 Int. Cl.:

H04S 3/00

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **22.11.2011 E 11793700 (3)**

97 Fecha y número de publicación de la concesión europea: **21.01.2015 EP 2649815**

54 Título: **Aparato y método para descomponer una señal de entrada utilizando una curva de referencia calculada previamente**

30 Prioridad:

10.12.2010 US 421927 P
11.05.2011 EP 11165746

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
20.04.2015

73 Titular/es:

**FRAUNHOFER-GESELLSCHAFT ZUR
FÖRDERUNG DER ANGEWANDTEN
FORSCHUNG E.V. (100.0%)**
Hansastraße 27c
80686 München, DE

72 Inventor/es:

WALTHER, ANDREAS

74 Agente/Representante:

PONTI SALES, Adelaida

ES 2 534 180 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Aparato y método para descomponer una señal de entrada utilizando una curva de referencia calculada previamente

5 **[0001]** La presente invención se refiere a procesamiento de audio y, en particular, a la descomposición de señales de audio en componentes diferentes tales como componentes distintos perceptualmente.

[0002] El sistema auditivo humano detecta sonido procedente de todas las direcciones. El entorno auditivo percibido (el adjetivo auditivo indica lo que se percibe, mientras que la palabra sonido se utilizará para describir fenómenos físicos) crea una impresión de las propiedades acústicas del espacio circundante y de los eventos sonoros que se producen. La impresión auditiva percibida en un campo de sonido específico se puede modelizar (por lo menos parcialmente) considerando tres tipos diferentes de señales en las entradas de los oídos: el sonido directo, las reflexiones tempranas y las reflexiones difusas. Estas señales contribuyen a la formación de una imagen espacial auditiva percibida.

15 **[0003]** El sonido directo indica las ondas de cada evento sonoro que llegan primero al oyente desde una fuente de sonido sin perturbaciones. Es característico de la fuente de sonido y proporciona la información menos comprometida sobre la dirección de incidencia del evento sonoro. Los indicios principales para estimar la dirección de una fuente de sonido en el plano horizontal son las diferencias entre las señales de entrada de los oídos izquierdo y derecho, es decir, las diferencias interaurales de tiempo (ITDs, interaural time differences) y las diferencias interaurales de nivel (ILDs, interaural level differences). A continuación, llegan a los oídos multitud de reflexiones del sonido directo desde direcciones diferentes y con diferentes niveles y retardos de tiempo relativos. A medida que aumenta el retardo temporal, en relación con el sonido directo, aumenta la densidad de las reflexiones hasta que éstas constituyen un artefacto estadístico.

25 **[0004]** El sonido reflejado contribuye a la percepción de la distancia, y a la impresión espacial auditiva, que se compone, por lo menos, de dos componentes: la anchura aparente de la fuente (ASW, apparent source width) (otro término utilizado normalmente para la ASW es amplitud auditiva) y la envoltura del oyente (LEV, listener envelopment). La ASW se define como un ensanchamiento de la anchura aparente de una fuente de sonido y está determinada principalmente por reflexiones laterales tempranas. La LEV se refiere a la sensación del oyente de estar envuelto por el sonido y está determinada principalmente por reflexiones que llegan tarde. El objetivo de la reproducción electroacústica estereofónica del sonido es evocar la percepción de una imagen espacial auditiva agradable. Esto puede tener una referencia natural o arquitectónica (por ejemplo, la grabación de un concierto en una sala), o puede ser un campo de sonido que no existe en la realidad (por ejemplo, música electroacústica).

35 **[0005]** En el sector de la acústica de salas de conciertos, es bien conocido que -para obtener un campo de sonido agradable subjetivamente- es importante una sensación fuerte de impresión espacial auditiva, siendo la LEV una parte integral. Es de interés la capacidad de las configuraciones de altavoces para reproducir un campo de sonido envolvente mediante la reproducción de un campo de sonido difuso. En un campo de sonido sintético no es posible reproducir utilizando transductores dedicados todas las reflexiones que se producen naturalmente. Éste es especialmente el caso para reflexiones tardías difusas. Las propiedades de temporización y de nivel de las reflexiones difusas se pueden simular utilizando señales "reverberadas" como entradas a los altavoces. Si éstas son lo suficientemente no correlacionadas, el número y la posición de los altavoces utilizados para la reproducción determina si el campo de sonido se percibe como siendo difuso. El objetivo es evocar la percepción de un campo de sonido continuo, difuso, utilizando solamente un número discreto de transductores. Es decir, crear campos de sonido donde no se pueda estimar ninguna dirección de llegada del sonido y no se pueda localizar en especial ningún transductor individual. El grado de difusión subjetivo de los campos de sonido sintéticos se puede evaluar en pruebas subjetivas.

50 **[0006]** Las reproducciones de sonido estereofónico tienen por objetivo evocar la percepción de un campo de sonido continuo utilizando solamente un número discreto de transductores. Las características más deseadas son la estabilidad direccional de fuentes localizadas y la representación realista del entorno auditivo circundante. La mayoría de los formatos utilizados actualmente para almacenar o transportar grabaciones estereofónicas están basados en canales. Cada canal transporta una señal que está prevista para ser reproducida en un altavoz asociado, en una posición específica. Una imagen auditiva específica se diseña durante el proceso de grabación o de mezcla. La imagen se recrea de manera precisa si la configuración de altavoces utilizada para la reproducción se parece a la configuración objetivo para la que se diseñó la grabación.

60 **[0007]** El número de canales factibles de transmisión y reproducción crece constantemente, y con cada formato emergente de reproducción de audio surge el deseo de representar contenido de formatos heredados sobre el sistema de reproducción actual. Los algoritmos de mezcla hacia arriba son una solución a este deseo, calculando una señal con más canales a partir de una señal heredada. En la bibliografía se ha propuesto un cierto número de algoritmos de mezcla hacia arriba en estéreo, por ejemplo, Carlos Avendano y Jean-Marc Jot, "A frequency-domain approach to multichannel upmix" (un enfoque en el dominio de frecuencia para una mezcla hacia arriba de múltiples canales), Journal of the Audio Engineering Society, volumen 52, número 7/8, páginas 740 a 749, 2004; Christof Faller, "Multiple-loudspeaker playback of stereo signals" (reproducción de señales estéreo en múltiples altavoces),

Journal of the Audio Engineering Society, volumen 54, número 11, páginas 1051 a 1064, noviembre de 2006; John Usherand Jacob Benesty, "Enhancement of spatial sound quality: A new reverberation-extraction audio upmixer" (realce de la calidad espacial del sonido: un nuevo mezclador hacia arriba de audio de reverberación-extracción), IEEE Transactions on Audio, Speech, and Language Processing, volumen 15, número 7, páginas 2141 a 2150, 5 septiembre de 2007. La mayor parte de estos algoritmos están basados en una descomposición de señales directa/ambiente seguida por una representación adaptada a la configuración de los altavoces objetivo.

[0008] Las descomposiciones de señales directa/ambiente descritas no son aplicables fácilmente a señales envolventes de múltiples canales. No es sencillo formular un modelo de señal y un filtrado para obtener a partir de N 10 canales de audio los correspondientes N canales de sonido directo y N canales de sonido ambiente. El modelo simple de señal utilizado en el caso estéreo, ver por ejemplo Christof Faller, "Multiple-loudspeaker playback of stereo signals", Journal of the Audio Engineering Society, volumen 54, número 11, páginas 1051 a 1064, noviembre de 2006, asumiendo que se va a correlacionar sonido directo entre todos los canales, no captura la diversidad de las relaciones de canales que existen entre canales de señales envolventes.

15 **[0009]** El objetivo general de la reproducción de sonido estereofónico es evocar la percepción de un campo de sonido continuo utilizando solamente un número limitado de canales de transmisión y transductores. Dos altavoces son el requisito mínimo para la reproducción espacial del sonido. Los sistemas modernos de los consumidores ofrecen a menudo un gran número de canales de reproducción. Básicamente, las señales estereofónicas 20 (independientemente del número de canales) se graban o se mezclan de manera que para cada fuente, el sonido directo va coherente (= dependiente) a una serie de canales con indicios direccionales específicos, y los sonidos independientes reflejados van a una serie de canales que determinan indicios para la anchura aparente de la fuente y la envoltura del oyente. Habitualmente, la percepción correcta de la imagen auditiva prevista es posible solamente en el punto de observación ideal en la configuración de la reproducción para la que se ha previsto la grabación. 25 Añadir más altavoces a una configuración de altavoces dada permite habitualmente una reconstrucción/simulación más realista del campo de sonido natural. Para utilizar todas las ventajas de la configuración extendida de altavoces si las señales de entrada están proporcionadas en otro formato, o para manipular las partes distintas perceptualmente de la señal de entrada, éstas tienen que ser accesibles por separado. Esta memoria descriptiva describe a continuación un método para separar las componentes dependientes e independientes de grabaciones 30 estereofónicas que comprenden un número arbitrario de canales de entrada.

[0010] Una descomposición de las señales de audio en componentes distintos perceptualmente es necesaria para modificación, realce, reproducción adaptativa y codificación perceptual de señales, de alta calidad. Recientemente se han propuesto una serie de métodos que permiten la manipulación y/o extracción de componentes de señal 35 distintos perceptualmente a partir de señales de entrada de dos canales. Dado que las señales de entrada con más de dos canales son cada vez más comunes, las manipulaciones descritas son deseables asimismo para señales de entrada de múltiples canales. Sin embargo, la mayor parte de los conceptos descritos para una entrada de dos canales no se pueden extender fácilmente para trabajar con señales de entrada con un número arbitrario de canales.

40 **[0011]** Si se fuera a realizar un análisis de señales en partes directa y ambiente, por ejemplo, con una señal envolvente de canales 5.1 que tiene un canal izquierdo, un canal central, un canal derecho, un canal envolvente izquierdo, un canal envolvente derecho y una realce de baja frecuencia (altavoz de graves), no está claro cómo se debería aplicar un análisis de señales directa/ambiente. Se puede podría considerar comparar cada par de seis 45 canales, lo que tiene como resultado un proceso jerárquico que tiene, al final, hasta 15 diferentes operaciones de comparación. A continuación, cuando se han realizado la totalidad de éstas 15 operaciones de comparación, donde cada canal se ha comparado con cada uno de todos los demás canales, habría que determinar cómo se deberían evaluar los 15 resultados. Esto consume mucho tiempo, los resultados son difíciles de interpretar, y debido a la cantidad considerable de recursos de procesamiento, no se puede utilizar, por ejemplo, para aplicaciones en tiempo 50 real de separación directo/ambiente o, en general, descomposiciones de señal que pueden ser utilizadas, por ejemplo, en el contexto de una mezcla hacia arriba o en cualesquiera otras operaciones de procesamiento de audio.

[0012] En el documento de M. M. Goodwin y J. M. Jot, "Primary-ambient signal decomposition and vector-based localization for spatial audio coding and enhancement" (descomposición de señales principal-ambiente y localización 55 basada en vectores para codificación de audio espacial y realce) en Proc. Of ICASSP 2007, 2007, se aplica un análisis de componentes principales a las señales del canal de entrada para realizar la descomposición de señal principal (= directa) y de ambiente.

[0013] Los modelos utilizados en el documento de Christof Faller, "Multiple-loudspeaker playback of stereo 60 signals", Journal of the Audio Engineering Society, volumen 54, número 11, páginas 1051 a 1064, noviembre de 2006 y en el documento de C Faller, "A highly directive 2-capsule based microphone system" (un sistema de micrófonos basado en 2 cápsulas altamente direccional) en Preprint 123rd Conv. Aud. Eng. Soc., octubre de 2007, asumen sonido difuso descorrelacionado o parcialmente correlacionado en señales estéreo y de micrófono, respectivamente. Con esta hipótesis, obtienen filtros para extraer la señal difusa/de ambiente. Estos enfoques están 65 limitados a señales de audio de un solo canal y de dos canales.

[0014] Una referencia adicional es C.Avendano y Jean-Marc Jot, "A frequency-domain approach to multichannel upmix", Journal of the Audio Engineering Society, volumen 52, número 7/8, páginas 740 a 749, 2004 La referencia de M. M. Goodwin y J. M. Jot, "Primary-ambient signal decomposition and vector-based localization for spatial audio coding and enhancement" en Proc. Of ICASSP 2007, 2007, comenta la referencia de Avendano, Jot como sigue. La referencia da a conocer un enfoque que involucra crear una máscara de tiempo-frecuencia para extraer el ambiente a partir de la señal de entrada estéreo. Sin embargo, la máscara está basada en la correlación cruzada entre las señales de canales izquierdo y derecho, de manera que este enfoque no es inmediatamente aplicable al problema de extraer ambiente a partir de una entrada arbitraria de múltiples canales. Utilizar cualquiera de dichos métodos basados en correlación, en este caso de orden superior, exigiría un análisis de correlación jerárquico por pares, que implicaría un coste computacional significativo, o alguna otra medida de correlación de múltiples canales.

[0015] La representación de respuesta de impulso espacial (SIRR, Spatial Impulse Response Rendering) (Juha Merimaa y Ville Pulkki, "Spatial impulse response rendering", en Proc. of the 7th Int. Conf. on Digital Audio Effects (DAFx'04), 2004) estima el sonido directo con dirección y el sonido difuso en respuestas de impulso en formato B. Muy similar a SIRR, la codificación de audio direccional (DirAC, Directional Audio Coding) (Ville Pulkki, "Spatial sound reproduction with directional audio coding" (reproducción espacial de sonido con codificación de audio direccional), Journal of the Audio Engineering Society, volumen 55, número 6, páginas 503 a 516, junio de 2007) implementa un análisis similar de sonido directo y difuso para señales de audio continuas de formato B.

[0016] El enfoque presentado en Jakka, Binaural to Multichannel Audio Upmix (mezcla hacia arriba de audio binaural de múltiples canales), tesis doctoral, tesis de maestría, Helsinki University of Technology, 2005, describe una mezcla hacia arriba utilizando señales binaurales como entrada.

[0017] La referencia de Boaz Rafaely, "Spatially Optimal Wiener Filtering in a Reverberant Sound Field, IEEE Workshop on Applications of Signal Processing to Audio and Acoustics 2001, October 21 to 24, 2001, New Paltz, New York", describe la obtención de filtros de Wiener óptimos espacialmente para campos de sonido reverberante. Se proporciona una aplicación para cancelación de ruido para dos micrófonos en habitaciones reverberantes. Los filtros óptimos que se obtienen de la correlación espacial de campos de sonido difusos capturan el comportamiento local de los campos de sonido y, por lo tanto, son de orden inferior y potencialmente más robustos espacialmente que los filtros de cancelación de ruido adaptativos convencionales en habitaciones reverberantes. Se presentan fórmulas para filtros óptimos no limitados y limitados causalmente, y se demuestra utilizando una simulación informática una aplicación de ejemplo para un realce de voz con dos micrófonos.

[0018] El documento WO 2009/100876 se considera la técnica anterior más próxima al tema de las reivindicaciones independientes 1, 14 y 15.

[0019] Este documento muestra un aparato para descomponer una señal que tiene una serie de canales, que comprende:
 un analizador para analizar una similitud entre dos canales de una señal de análisis en relación con la señal que tiene por lo menos dos canales de análisis, donde el analizador está configurado para utilizar una curva de referencia a efectos de determinar el resultado del análisis; y
 un procesador de señal para procesar la señal de análisis o una señal obtenida a partir de la señal de análisis, o una señal de la que se deriva la señal de análisis, utilizando el resultado del análisis para obtener una señal descompuesta.

[0020] Si bien el enfoque de filtro de Wiener puede proporcionar resultados útiles para la cancelación del ruido en habitaciones reverberantes, puede ser ineficiente computacionalmente y, en algunos casos, no es muy útil para la descomposición de señales.

[0021] El objetivo de la presente invención es dar a conocer un concepto mejorado para la descomposición de una señal de entrada.

[0022] Este objetivo se consigue mediante un aparato para descomponer una señal de entrada según la reivindicación 1, un método para descomponer una señal de entrada según la reivindicación 14 o un programa informático según la reivindicación 15.

[0023] La presente invención está basada en el hallazgo de que se obtiene una eficiencia particular para el propósito de descomposición de señales cuando el análisis de las señales se realiza en base a la curva de similitud dependiente de la frecuencia calculada previamente, como una curva de referencia. El término similitud incluye la correlación y la coherencia, donde en sentido matemático -estricto-, la correlación se calcula entre dos señales sin un desplazamiento de tiempo adicional y la coherencia se calcula desplazando las dos señales en tiempo/fase, de manera que las señales tienen una correlación máxima, y a continuación la correlación real sobre la frecuencia se calcula con el desplazamiento de tiempo/fase aplicado. Para este texto, se considera que similitud, correlación y coherencia significan lo mismo, es decir, el grado cuantitativo de similitud entre dos señales, por ejemplo, donde un mayor valor absoluto de la similitud significa que las dos señales son más similares y un menor valor absoluto de la similitud significa que las dos señales son menos similares.

[0024] Se ha demostrado que la utilización de dicha curva de similitud como curva de referencia permite un análisis implementable de manera muy eficiente, dado que la curva se puede utilizar para operaciones de comparación directa y/o cálculos de factores de ponderación. La utilización de una curva de similitud dependiente de la frecuencia calculada previamente permite limitarse a realizar cálculos simples en lugar de operaciones más complejas de filtro de Wiener. Además, la aplicación de la curva de similitud dependiente de la frecuencia es particularmente útil debido al hecho de que el problema no se trata desde un punto de vista estadístico sino que se trata de manera más analítica, dado que se introduce toda la información posible a partir de la configuración actual para obtener una solución al problema. Adicionalmente, la flexibilidad de este procedimiento es muy elevada, dado que la curva de referencia se puede obtener de muchas maneras diferentes. Una manera es medir realmente las dos o más señales en una cierta configuración, y calcular a continuación la curva de similitud sobre la frecuencia a partir de las señales medidas. Por lo tanto, se pueden emitir señales independientes desde altavoces diferentes, o señales con un cierto grado de dependencia que se conoce por adelantado.

[0025] La otra alternativa preferida es calcular simplemente la curva de similitud suponiendo señales independientes. En este caso, realmente no es necesaria ninguna señal, dado que el resultado es independiente de las señales.

[0026] La descomposición de señales utilizando una curva de referencia para el análisis de señales se puede aplicar para procesamiento estéreo, es decir, para descomponer una señal estéreo. Alternativamente, este procedimiento se puede implementar asimismo junto con un mezclador hacia abajo para descomponer señales de múltiples canales. Alternativamente, este procedimiento se puede implementar asimismo para señales de múltiples canales sin utilizar un mezclador hacia abajo cuando se contempla una evaluación de señales por pares, de manera jerárquica.

[0027] En otra realización, un enfoque ventajoso consiste en no realizar el análisis con respecto a los diferentes componentes de señal con la señal de entrada directamente, es decir con una señal que tiene por lo menos tres canales de entrada. En lugar de esto, la señal de entrada de múltiples canales que tiene por lo menos tres canales de entrada se procesa mediante un mezclador hacia abajo para mezclar hacia abajo la señal de entrada a efectos de obtener una señal mezclada hacia abajo. La señal mezclada hacia abajo tiene una serie de canales de mezcla hacia abajo, que es menor que el número de canales de entrada y, preferentemente, es de dos. A continuación, el análisis de la señal de entrada se realiza sobre la señal mezclada hacia abajo en lugar de sobre la señal de entrada directamente, y el análisis tiene como resultado un resultado del análisis. Sin embargo, este resultado del análisis no se aplica a la señal mezclada hacia abajo, sino que se aplica a la señal de entrada o, alternativamente, a una señal obtenida a partir de la señal de entrada, donde esta señal obtenida a partir de la señal de entrada puede ser una señal de mezcla hacia arriba o, dependiendo del número de canales de las señales de entrada, puede ser asimismo una señal de mezcla hacia abajo, pero esta señal obtenida a partir de la señal de entrada será diferente a la señal mezclada hacia abajo, sobre la que se ha realizado el análisis. Por ejemplo, cuando se considera el caso de que la señal de entrada es una señal de canal 5.1, entonces la señal de mezcla hacia abajo, sobre la que se realiza el análisis, puede ser una mezcla hacia abajo estéreo con dos canales. A continuación, los resultados del análisis se aplican directamente a la señal de entrada 5.1, a una mezcla hacia arriba superior tal como una señal de salida 7.1 o a una mezcla hacia abajo de múltiples canales de la señal de entrada que tiene por ejemplo solamente tres canales, que son el canal izquierdo, el canal central y el canal derecho, donde solamente está disponible un aparato de representación de audio de tres canales. Sin embargo, en cualquier caso, la señal sobre la que se aplican los resultados del análisis mediante el procesador de señal es diferente a la señal mezclada hacia abajo sobre la que se ha realizado el análisis, y habitualmente tiene más canales que la señal mezclada hacia abajo, sobre la que se ha realizado el análisis con respecto a los componentes de la señal.

[0028] El denominado análisis/procesamiento "indirecto" es posible debido al hecho de que se puede asumir que cualesquiera componentes en los canales de entrada individuales se pueden producir asimismo en los canales mezclados hacia abajo, dado que la mezcla hacia abajo consiste habitualmente en una suma de canales de entrada de diferentes maneras. Una mezcla hacia abajo directa es, por ejemplo, que los canales de entrada individuales se ponderen según proceda mediante una regla de mezcla hacia abajo o una matriz de mezcla hacia abajo, y a continuación se sumen entre sí después de haber sido ponderados. Una mezcla hacia abajo alternativa consiste en filtrar los canales de entrada con ciertos filtros, tales como filtros HRTF, y la mezcla hacia abajo se realiza utilizando señales filtradas, es decir, las señales filtradas por filtros HRTF, tal como se conoce en la técnica. Para una señal de entrada de cinco canales se requieren 10 filtros HRTF, y las salidas de los filtros HRTF para la parte izquierda/oido izquierdo se suman entre sí y las salidas de los filtros HRTF para los filtros del canal derecho se suman entre sí para el oído derecho. Se pueden aplicar mezclas hacia abajo alternativas para reducir el número de canales que hay que procesar en el analizador de señal.

[0029] Por lo tanto, las realizaciones de la presente invención describen un concepto nuevo para extraer componentes distintos perceptualmente a partir de señales de entrada arbitrarias, considerando una señal de análisis, donde el resultado del análisis se aplica a la señal de entrada. Dicha señal de análisis se puede adquirir, por ejemplo, considerando un modelo de propagación de los canales o de las señales de los altavoces hacia los oídos. Esto está motivado, en parte, por el hecho de que el sistema auditivo humano utiliza asimismo solamente dos

sensores (el oído izquierdo y el derecho) para evaluar campos de sonido. Por lo tanto, la extracción de componentes distintos perceptualmente se reduce básicamente a la consideración de una señal de análisis que se indicará como mezcla hacia abajo en lo que sigue. En todo este documento, el término mezcla hacia abajo se utiliza para cualquier procesamiento previo de la señal de múltiples canales que tiene como resultado una señal de análisis (esto puede incluir, por ejemplo, un modelo de propagación, HRTFs, BRIRs, mezcla hacia abajo simple de factores cruzados).

[0030] Conociendo el formato de la entrada dada y las características deseadas de la señal a extraer, se pueden definir las relaciones ideales entre canales para el formato mezclado hacia abajo y dicho análisis de esta señal de análisis es suficiente para generar una máscara de ponderación (o múltiples máscaras de ponderación) para la descomposición de señales de múltiples canales.

[0031] En una realización, el problema de múltiples canales se simplifica utilizando una mezcla hacia abajo estéreo de una señal envolvente y aplicando un análisis directo/ambiente a la mezcla hacia abajo. En base a este resultado, es decir, a estimaciones de espectros de potencia de tiempo reducido de sonidos directo y de ambiente, se obtienen filtros para descomponer una señal de N canales en N canales de sonido directo y N canales de sonido ambiente.

[0032] La presente invención es ventajosa debido al hecho de que se aplica un análisis de señales sobre un número menor de canales, lo que reduce significativamente el tiempo necesario de procesamiento, de manera que el concepto inventivo se puede aplicar incluso en aplicaciones en tiempo real para mezcla hacia arriba o mezcla hacia abajo, o para cualquier otra operación de procesamiento de señales donde se requieran componentes diferentes, tales como componentes perceptualmente diferentes, de una señal.

[0033] Una ventaja adicional de la presente invención es que, aunque se realiza una mezcla hacia abajo, se ha encontrado que esto no deteriora la detectabilidad de componentes distintos perceptualmente en la señal de entrada. Dicho de otra manera, incluso cuando se mezclan hacia abajo canales de entrada, los componentes de señal individuales pueden sin embargo separarse en gran medida. Además, la mezcla hacia abajo funciona como una clase de "recopilación" de todos los componentes de señal de todos los canales de entrada en dos canales, y el único análisis aplicado sobre estas señales mezcladas hacia abajo "recopiladas" proporciona un resultado único que ya no tiene que ser interpretado y que se puede utilizar directamente para el procesamiento de señales.

[0034] A continuación se comentan realizaciones preferidas de la presente invención, con respecto a las figuras adjuntas, en las cuales:

la figura 1 es un diagrama de bloques para mostrar un aparato para descomponer una señal de entrada utilizando un mezclador hacia abajo;

la figura 2 es un diagrama de bloques que muestra una implementación de un aparato para descomponer una señal que tiene una serie de por lo menos tres canales de entrada, utilizando un analizador con una curva de correlación calculada previamente dependiente de la frecuencia, de acuerdo con otro aspecto de la invención;

la figura 3 muestra otra implementación preferida de la presente invención con un procesamiento en el dominio de frecuencia para la mezcla hacia abajo, el análisis y el procesamiento de señal;

la figura 4 muestra una curva de correlación calculada previamente dependiente de la frecuencia, a modo de ejemplo, para una curva de referencia para el análisis indicado en la figura 1 o la figura 2;

la figura 5 muestra un diagrama de bloques que ilustra otro proceso para extraer componentes independientes;

la figura 6 muestra otra implementación de un diagrama de bloques para un procesamiento posterior en el que se extraen componentes difuso independiente, directo independiente y directo;

la figura 7 muestra un diagrama de bloques que implementa el mezclador hacia abajo como un generador de señal de análisis;

la figura 8 muestra un diagrama de flujo para indicar una manera preferida de procesamiento en el analizador de señal de la figura 1 o la figura 2;

las figuras 9a a 9e muestran diferentes curvas de correlación calculadas previamente dependientes de la frecuencia, que se pueden utilizar como curvas de referencia para varias configuraciones diferentes, con diferentes cantidades y posiciones de fuentes de sonido (tales como altavoces);

la figura 10 muestra un diagrama de bloques para ilustrar otra realización para la estimación del grado de difusión, donde los componentes difusos son los componentes a descomponer; y

las figuras 11A y 11B muestran ecuaciones a modo de ejemplo, para aplicar un análisis de señal sin una curva de correlación dependiente de la frecuencia, sino dependiendo del enfoque de filtrado de Wiener.

[0035] La figura 1 muestra un aparato para descomponer una señal de entrada 10 que tiene una serie de, por lo menos, tres canales de entrada o, en general, N canales de entrada. Estos canales de entrada se introducen en un mezclador hacia abajo 12 para mezclar hacia abajo la señal de entrada a efectos de obtener una señal mezclada hacia abajo 14, donde el mezclador hacia abajo 12 está dispuesto para realizar una mezcla hacia abajo de tal modo que el número de canales de mezcla hacia abajo de la señal mezclada hacia abajo 14, que se indica mediante "m", es de por lo menos dos y es menor que el número de canales de entrada de la señal de entrada 10. Los m canales de mezcla hacia abajo se introducen en el analizador 16 para analizar la señal mezclada hacia abajo con el fin de obtener un resultado del análisis 18. El resultado del análisis 18 se introduce en un procesador de señal 20, donde el procesador de señal está dispuesto para procesar la señal de entrada 10 o una señal obtenida a partir de la señal de entrada mediante un dispositivo 22 de obtención de señales utilizando el resultado del análisis, donde el procesador de señal 20 está configurado para aplicar los resultados del análisis a los canales de entrada o a canales de la señal 24 obtenidos a partir de la señal de entrada, a efectos de obtener una señal descompuesta 26.

[0036] En la realización mostrada en la figura 1, el número de canales de entrada es de n, el número de canales de mezcla hacia abajo es de m, el número de canales obtenidos es de l, y el número de canales de salida es igual a l, donde se procesa mediante el procesador de señal la señal obtenida en lugar de la señal de entrada. Alternativamente, cuando el dispositivo 22 de obtención de señales no existe, entonces la señal de entrada es procesada directamente por el procesador de señal y entonces el número de canales de la señal descompuesta 26, indicado mediante "l" en la figura 1, será igual a n. Por lo tanto, la figura 1 muestra dos ejemplos diferentes. Un ejemplo no tiene el dispositivo 22 de obtención de señales y la señal de entrada se aplica directamente al procesador de señal 20. El otro ejemplo consiste en que está implementado el dispositivo 22 de obtención de señales y, entonces, se procesa mediante el procesador de señal 20 la señal obtenida 24 en lugar de la señal de entrada 10. El dispositivo de obtención de señales puede ser, por ejemplo, un mezclador de canales de audio tal como un mezclador hacia arriba, para generar más canales de salida. En este caso, l sería mayor que n. En otra realización, el dispositivo de obtención de señales podría ser otro procesador de audio que aplique ponderación, retardo o cualquier otra cosa sobre los canales de entrada y, en este caso, el número de canales de salida l del dispositivo 22 de obtención de señales sería igual al número n de canales de entrada. En otra implementación, el dispositivo de obtención de señales sería un mezclador hacia abajo que reduce el número de canales desde la señal de entrada a la señal obtenida. En esta implementación, es preferible que el número l siga siendo mayor que el número m de canales mezclados hacia abajo, para tener una de las ventajas de la presente invención, es decir, que el análisis de señal se aplica a un número menor de señales de canal. El analizador es operativo para analizar la señal mezclada hacia abajo, con respecto a componentes distintos perceptualmente. Estos componentes distintos perceptualmente pueden ser componentes independientes en los canales individuales por una parte, y componentes dependientes por otra. Los componentes de señal alternativos a analizar mediante la presente invención son componentes directos por una parte y componentes ambiente por otra parte. Existen muchos otros componentes que se pueden separar mediante la presente invención, tales como componentes de voz respecto de componentes de música, componentes de ruido respecto de componentes de voz, componentes de ruido respecto de componentes de música, componentes de ruido de alta frecuencia respecto de componentes de ruido de baja frecuencia, en señales de múltiples alturas, los componentes proporcionadas por los diferentes instrumentos, etc. Esto se debe al hecho de que existen potentes herramientas de análisis, tales como el filtrado de Wiener que se ha comentado en el contexto de las figuras 11A, 11B u otros procedimientos de análisis tales como utilizar una curva de correlación dependiente de la frecuencia que se ha comentado, por ejemplo, en el contexto de la figura 8 según la presente invención.

[0037] La figura 2 muestra otro aspecto, donde el analizador está implementado para utilizar una curva de correlación calculada previamente dependiente de la frecuencia 16. Por lo tanto, el aparato para descomponer una señal 28 que tiene una serie de canales comprende el analizador 16 para analizar una correlación entre dos canales, de una señal de análisis idéntica a la señal de entrada o relacionada con la señal de entrada, por ejemplo, mediante una operación de mezcla hacia abajo, tal como se muestra en el contexto de la figura 1. La señal de análisis analizada por el analizador 16 tiene, por lo menos, dos canales de análisis, y el analizador 16 está configurado para utilizar una curva de correlación calculada previamente dependiente de la frecuencia, como una curva de referencia para determinar el resultado del análisis 18. El procesador de señal 20 puede funcionar del mismo modo que se ha comentado en el contexto de la figura 1, y está configurado para procesar la señal de análisis o una señal obtenida a través de la señal de análisis mediante un dispositivo 22 de obtención de señales, donde el dispositivo 22 de obtención de señales se puede implementar de manera similar a lo que se ha comentado en el contexto del dispositivo 22 de obtención de señales de la figura 1. Alternativamente, el procesador de señal puede procesar una señal, a partir de la cual se obtiene la señal de análisis y el procesamiento de señal utiliza el resultado del análisis para obtener una señal descompuesta. Por lo tanto, en la realización de la figura 2, la señal de entrada puede ser idéntica a la señal de análisis y, en este caso, la señal de análisis puede ser asimismo una señal estéreo que tiene solamente dos canales, tal como se muestra en la figura 2. Alternativamente, la señal de análisis se puede obtener a partir de una señal de entrada mediante cualquier clase de proceso, tal como mezcla hacia abajo, según se describe en el contexto de la figura 1, o mediante cualquier otro proceso tal como mezcla hacia arriba o similares. Adicionalmente, el procesador de señal 20 puede ser útil para aplicar el procesamiento de señal a la misma señal que se ha introducido en el analizador, o el procesador de señal puede aplicar un procesamiento de señal a una señal a partir de la cual se ha obtenido la señal de análisis, tal como se ha indicado en el contexto de la figura 1, o el

procesador de señal puede aplicar un procesamiento de señal a una señal que se ha obtenido a partir de la señal de análisis, tal como mediante mezcla hacia arriba o similares.

[0038] Por lo tanto, existen diferentes posibilidades para el procesador de señal, y todas estas posibilidades son ventajosas debido a la operación única del analizador utilizando una curva de correlación calculada previamente dependiente de la frecuencia como curva de referencia, para determinar el resultado del análisis.

[0039] A continuación se comentan otras realizaciones. Se debe observar que, tal como se ha comentado en el contexto de la figura 2, se considera incluso la utilización de una señal de análisis de dos canales (sin una mezcla hacia abajo). Por lo tanto, en la presente invención tal como se comenta en los diferentes aspectos en el contexto de la figura 1 y la figura 2, que se pueden utilizar juntos o como aspectos independientes, la mezcla hacia abajo puede ser procesada por el analizador o una señal de dos canales, que probablemente no ha sido generada mediante una mezcla hacia abajo, puede ser procesada mediante el analizador de señal utilizando la curva de referencia calculada previamente. En este contexto, se debe observar que la descripción posterior de aspectos de la implementación se puede aplicar a ambos aspectos mostrados esquemáticamente en la figura 1 y en la figura 2, incluso cuando ciertas características estén descritas solamente para un aspecto y no para ambos. Por ejemplo, si se considera la figura 3, resulta evidente que las características del dominio de frecuencia de la figura 3 se describen en el contexto del aspecto mostrado en la figura 1, pero resulta evidente que una transformada de tiempo/frecuencia, tal como se describe a continuación con respecto a la figura 3, y la transformada inversa, se pueden aplicar asimismo a la implementación de la figura 2, que no tiene un mezclador hacia abajo, pero que tiene un analizador específico que utiliza una curva de correlación calculada previamente dependiente de la frecuencia.

[0040] En particular, el convertidor de tiempo/frecuencia estaría situado para convertir la señal de análisis antes de que la señal de análisis se introduzca en el analizador, y se situaría asimismo el convertidor de frecuencia/tiempo a la salida del procesador de señal para convertir la señal procesada de nuevo al dominio de tiempo. Cuando existe un dispositivo de obtención de señales, el convertidor de tiempo/frecuencia se puede colocar a la entrada del dispositivo de obtención de señales, de tal modo que el dispositivo de obtención de señales, el analizador y el procesador de señal funcionan todos en el dominio de frecuencia/sub-banda. En este contexto, frecuencia y sub-banda significan básicamente una parte de la frecuencia en una representación de frecuencias.

[0041] Además, resulta evidente que el analizador de la figura 1 se puede implementar de muchas maneras diferentes, pero éste analizador está implementado asimismo, en una realización, como el analizador comentado en la figura 2, es decir como un analizador que utiliza una curva de correlación calculada previamente dependiente de la frecuencia como alternativa al filtro de Wiener o a cualquier otro método de análisis.

[0042] La realización de la figura 3 aplica un procedimiento de mezcla hacia abajo a una señal de entrada arbitraria a efectos de obtener una representación de dos canales. Se realiza un análisis del dominio de tiempo-frecuencia y se calculan máscaras de ponderación que se multiplican por la representación de tiempo-frecuencia de la señal de entrada, tal como se muestra en la figura 3.

[0043] En la imagen, T/F indica una transformada de tiempo-frecuencia; normalmente una transformada de Fourier de tiempo reducido (STFT, Short-time Fourier Transform). iT/F indica la transformada inversa respectiva. $[x_1(n), \dots, x_N(n)]$ son las señales de entrada, donde n es el índice de tiempo. $[X_1(m, i), \dots, X_N(m, i)]$ indican los coeficientes de descomposición de frecuencia, donde m es el índice de tiempo de la descomposición e i es el índice de frecuencia de la descomposición. $[D_1(m, i), D_2(m, i)]$ son los dos canales de la señal mezclada hacia abajo.

$$\begin{pmatrix} D_1(m, i) \\ D_2(m, i) \end{pmatrix} = \begin{pmatrix} H_{11}(i) & H_{12}(i) & \dots & H_{1N}(i) \\ H_{21}(i) & H_{22}(i) & \dots & H_{2N}(i) \end{pmatrix} \begin{pmatrix} X_1(m, i) \\ X_2(m, i) \\ \vdots \\ X_N(m, i) \end{pmatrix} \quad (1)$$

[0044] $W(m, i)$ es la ponderación calculada. $[Y_1(m, i), \dots, Y_N(m, i)]$ son las descomposiciones de frecuencia ponderadas de cada canal. $H_{ij}(i)$ son los coeficientes de mezcla hacia abajo, que pueden ser de valor real o de valor complejo, y los coeficientes pueden ser constantes en el tiempo o variables en el tiempo. Por lo tanto, los coeficientes de la mezcla hacia abajo puede ser constantes, o filtros tales como filtros HRTF, filtros de reverberación o filtros similares.

$$Y_j(m, i) = W_j(m, i) \cdot X_j(m, i), \text{ where } j = (1, 2, \dots, N) \quad (2)$$

[0045] En la figura 3, se representa el caso de aplicación de la misma ponderación a todos los canales.

$$Y_j(m,i) = W(m,i) \cdot X_j(m,i) \quad (3)$$

[$y_1(n), \dots, y_N(n)$] son las señales de salida en el dominio de tiempo que comprenden los componentes de la señal extraída. (La señal de entrada puede tener un número arbitrario de canales (N), producidos para una configuración arbitraria de altavoces de reproducción objetivo. La mezcla hacia abajo puede incluir HRTFs para obtener señales de entrada al oído, simulación de filtros auditivos, etc. La mezcla hacia abajo se puede llevar a cabo asimismo en el dominio de tiempo.)

[0046] En una realización, se calcula la diferencia entre una correlación de referencia (en todo este texto, el término correlación se utiliza como sinónimo de similitud entre canales y por lo tanto puede incluir asimismo evaluaciones de desplazamientos de tiempo, para lo que normalmente se utiliza el término coherencia. Incluso si se evalúan desplazamientos temporales, el valor resultante puede tener signo. Normalmente, la coherencia se define teniendo solamente valores positivos) como una función de la frecuencia ($C_{ref}(\omega)$), y la correlación real de la señal de entrada mezclada hacia abajo ($C_{sig}(\omega)$). Dependiendo de la desviación de la curva real respecto de la curva de referencia, se calcula un factor de ponderación para cada tesela de tiempo-frecuencia, indicando si comprende componentes dependientes o independientes. La ponderación de tiempo-frecuencia obtenida indica los componentes independientes y puede ya ser aplicada a cada canal de la señal de entrada para producir una señal de múltiples canales (número de canales igual al número de canales de entrada) que incluye partes independientes que se pueden percibir como distintas o difusas.

20

[0047] La curva de referencia se puede definir de diferentes maneras. Son ejemplos:

- Curva de referencia teórica ideal, para un campo de sonido difuso idealizado bidimensional o tridimensional compuesto de componentes independientes.
- 25 • La curva ideal que se puede conseguir con la configuración de altavoces objetivo de referencia para la señal de entrada dada (por ejemplo, configuración estéreo estándar con ángulos de acimut ($\pm 30^\circ$), o configuración de cinco canales estándar según ITU-R BS.775 con ángulos de acimut ($0^\circ, \pm 30^\circ, \pm 110^\circ$)).
- La curva ideal para la configuración de altavoces presente realmente (las posiciones reales se podrían medir o conocerse mediante entrada del usuario. La curva de referencia se puede calcular asumiendo reproducción de señales independientes sobre los altavoces dados).
- 30 • La potencia real de tiempo reducido dependiente de la frecuencia de cada canal de entrada se puede incorporar al cálculo de la referencia.

[0048] Dada una curva de referencia dependiente de la frecuencia ($C_{ref}(\omega)$), se puede definir un umbral superior ($C_{hi}(\omega)$) y un umbral inferior ($C_{lo}(\omega)$) (véase la figura 4). Las curvas umbral pueden coincidir con la curva de referencia ($C_{ref}(\omega) = C_{hi}(\omega) = C_{lo}(\omega)$), o definirse asumiendo umbrales de detectabilidad, o se pueden obtener de manera heurística.

[0049] Si la desviación de la curva real respecto de la curva de referencia está dentro de los límites proporcionados por los umbrales, el sector actual obtiene una ponderación que indica componentes independientes. Por encima del umbral superior o por debajo del umbral inferior, el sector se indica como dependiente. Esta indicación puede ser binaria, o gradual (es decir, siguiendo una función de decisión programada). En particular, si los umbrales superior e inferior coinciden con la curva de referencia, la ponderación aplicada está relacionada directamente con la desviación respecto de la curva de referencia.

45

[0050] Haciendo referencia a la figura 3, el numeral de referencia 32 muestra un convertidor de tiempo/frecuencia que se puede implementar como una transformada de Fourier de tiempo reducido, o como cualquier clase de banco de filtros que genera señales de sub-banda, tal como un banco de filtros QMF o similares. Independientemente de la implementación detallada del convertidor de tiempo/frecuencia 32, la salida del convertidor de tiempo/frecuencia es, para cada canal de entrada x_i , un espectro para cada periodo de tiempo de la señal de entrada. De este modo, el procesador de tiempo/frecuencia 32 se puede implementar para tomar siempre un bloque de muestras de entrada de una señal de canal individual, y para calcular la representación de frecuencias tal como un espectro de FFT que tiene líneas espectrales que se extienden desde una frecuencia inferior hasta una frecuencia superior. A continuación, para un siguiente bloque de tiempo, se realiza el mismo procedimiento de manera que, al final, se calcula una secuencia de espectros de tiempo reducido para cada señal de canal de entrada. Se dice que un cierto intervalo de frecuencia de un cierto espectro relacionado con un cierto bloque de muestras de entrada de un canal de entrada es una "tesela de tiempo/frecuencia" y, preferentemente, el análisis en el analizador 16 se realiza en base a estas teselas de tiempo/frecuencia. Por lo tanto, el analizador recibe, como entrada para una tesela de tiempo/frecuencia, el valor espectral a una primera frecuencia para un cierto bloque de muestras de entrada del primer canal de mezcla hacia abajo D_1 , y recibe el valor para la misma frecuencia y el mismo bloque (de tiempo) del segundo canal de mezcla hacia abajo D_2 .

[0051] Entonces, tal como se muestra por ejemplo en la figura 8, el analizador 16 está configurado para determinar (80) un valor de correlación entre los dos canales de entrada por cada sub-banda y bloque de tiempo, es decir, un

valor de correlación para una tesela de tiempo/frecuencia. A continuación, el analizador 16 recupera, en la realización mostrada con respecto a la figura 2 o a la figura 4, un valor de correlación (82) para la sub-banda correspondiente a partir de la curva de correlación de referencia. Cuando, por ejemplo, la sub-banda es la sub-banda indicada en 40 en la figura 4, entonces la etapa 82 tiene como resultado que el valor 41 indica una correlación entre -1 y +1, y el valor 41 es entonces el valor de correlación recuperado. A continuación, en la etapa 83, el resultado para la sub-banda utilizando el valor de correlación determinado a partir de la etapa 80 y el valor de correlación recuperado 41 obtenido en la etapa 82, se obtiene realizando una comparación y la posterior decisión, o se obtiene calculando una diferencia real. El resultado puede ser, tal como se ha comentado anteriormente, un resultado binario que indica que la tesela de tiempo/frecuencia actual considerada en la mezcla hacia abajo/señal de análisis tiene componentes independientes. Esta decisión se adoptará cuando el valor de correlación determinado realmente (en la etapa 80) sea igual al valor de correlación de referencia o esté muy cerca del valor de correlación de referencia.

[0052] Sin embargo, cuando se determine que el valor de correlación determinado indica una correlación absoluta mayor que el valor de correlación de referencia, se determina entonces que la tesela de tiempo/frecuencia considerada comprende componentes dependientes. De este modo, cuando la correlación de la tesela de tiempo/frecuencia de la mezcla hacia abajo o de la señal de análisis indica un valor de correlación absoluto mayor que la curva de referencia, se puede decir entonces que los componentes en esta tesela de tiempo/frecuencia son dependientes entre sí. Sin embargo, cuando se indica que el valor de correlación está muy próximo a la curva de referencia, se puede decir entonces que los componentes son independientes. Los componentes dependientes pueden recibir un primer valor de ponderación tal como 1, y los componentes independientes pueden recibir un segundo valor de ponderación tal como 0. Preferentemente, tal como se muestra en la figura 4, se utilizan umbrales superior e inferior que están separados de la línea de referencia, para proporcionar un resultado mejor, que es más adecuado que utilizar solamente la curva de referencia.

[0053] Además, con respecto a la figura 4, se debe observar que la correlación puede variar entre -1 y +1. Una correlación con signo negativo indica adicionalmente un desplazamiento de fase de 180° entre las señales. Por lo tanto, se podrían aplicar asimismo otras correlaciones que se extienden solamente entre 0 y 1, en las que la parte negativa de la correlación simplemente se hace positiva. En este procedimiento, se ignoraría entonces un desplazamiento de tiempo o un desplazamiento de fase con el propósito de la determinación de la correlación.

[0054] La manera alternativa de calcular el resultado, es calcular realmente la distancia entre el valor de correlación determinado en el bloque 80 y el valor de correlación recuperado, obtenido en el bloque 82, y determinar a continuación una métrica entre 0 y 1 como un factor de ponderación basado en la distancia. Mientras que la primera alternativa (1) de la figura 8 tiene como resultado solamente los valores de 0 ó 1, la posibilidad (2) tiene como resultado valores entre 0 y 1 y, en algunas implementaciones, estos son preferibles.

[0055] El procesador de señal 20 de la figura 3 se muestra como multiplicadores, y los resultados del análisis son un factor de ponderación determinado que se envía desde el analizador al procesador de señal, tal como se muestra en 84 en la figura 8, y a continuación se aplica a la correspondiente tesela de tiempo/frecuencia de la señal de entrada 10. Cuando, por ejemplo, el espectro realmente considerado es el espectro 20-ésimo en la secuencia de espectros, y cuando el sector de frecuencia realmente considerado es el 5-ésimo sector de frecuencia de este espectro 20-ésimo, entonces la tesela de tiempo/frecuencia se puede indicar como (20, 5), donde el primer número indica el número del bloque de tiempo y el segundo número indica el sector de frecuencia en este espectro. A continuación, el resultado del análisis para la tesela de tiempo/frecuencia (20, 5) se aplica a la correspondiente tesela de tiempo/frecuencia (20, 5) para cada canal de la señal de entrada en la figura 3 o, cuando está implementado un dispositivo de obtención de señales tal como se muestra en la figura 1, a la correspondiente tesela de tiempo/frecuencia de cada canal de la señal obtenida.

[0056] A continuación, se comenta en mayor detalle el cálculo de una curva de referencia. Sin embargo, para la presente invención, básicamente carece de importancia cómo se ha obtenido la curva de referencia. Puede ser una curva arbitraria o, por ejemplo, valores en una tabla de consulta que indican una relación ideal o deseada de las señales de entrada x_i en la señal D de mezcla hacia abajo o, y en el contexto de la figura 2, en la señal de análisis. La siguiente deducción es a modo de ejemplo.

[0057] La difusión física de un campo de sonido se puede evaluar mediante un método introducido por Cook et al (Richard K. Cook, R. V. Waterhouse, R.D. Berendt, Seymour Edelman, y Jr. M.C. Thompson, "Measurement of correlation coefficients in reverberant sound fields" (medición de coeficientes de correlación en campos de sonido reverberantes), Journal Of The Acoustical Society Of America, volumen 27, número 6, páginas 1072 a 1077, noviembre de 1955), utilizando el coeficiente de correlación (r) de la presión del sonido en régimen estacionario de ondas planas en dos puntos separados espacialmente, tal como se muestra en la siguiente ecuación (4).

$$r = \frac{\langle p_1(n) \cdot p_2(n) \rangle}{[\langle p_1^2(n) \rangle \cdot \langle p_2^2(n) \rangle]^{\frac{1}{2}}} \quad (4)$$

donde $p_1(n)$ y $p_2(n)$ son mediciones de la presión del sonido en dos puntos, n es el índice de tiempo y $\langle \cdot \rangle$ indica promedio en el tiempo. En un campo de sonido en régimen estacionario, se pueden obtener las relaciones siguientes:

$$r(k, d) = \frac{\text{seno}(kd)}{kd} \text{ (para campos de sonido tridimensionales), y} \quad (5)$$

$$r(k, d) = J_0(kd) \text{ (para campos de sonido bidimensionales)} \quad (6)$$

10 donde d es la distancia entre los dos puntos de medición y $k = \frac{2\pi}{\lambda}$ es el número de onda, siendo λ la longitud de onda. (La curva de referencia física $r(k, d)$ puede ya ser utilizada como C_{ref} para el procesamiento posterior.)

[0058] Una medida para el grado de difusión perceptual de un campo de sonido es el coeficiente de correlación cruzada interaural (ρ), medido en un campo de sonido. Medir ρ implica que el radio entre los sensores de presión (respectivamente, los oídos) es fijo. Incluyendo esta restricción, r pasa a ser función de la frecuencia, con la frecuencia angular $\omega = kc$, donde c es la velocidad del sonido en el aire. Además, las señales de presión difieren respecto de las señales en campo libre consideradas anteriormente, debido a la reflexión, la difracción y efectos de curvatura provocados por los pabellones auditivos, la cabeza y el torso del oyente. Dichos efectos, sustanciales para la audición espacial, están descritos mediante funciones de transferencia relativas a la cabeza (HRTFs, head-related transfer functions). Considerando estas influencias, las señales de presión resultantes en las entradas de los oídos son $p_L(n, \omega)$ y $p_R(n, \omega)$. Para el cálculo, se pueden utilizar datos HRTF medidos o se pueden obtener aproximaciones utilizando un modelo analítico (por ejemplo, Richard O. Duda y William L. Martens, "Range dependence of the response of a spherical head model" (dependencia de intervalos de la respuesta de un modelo esférico de la cabeza), Journal Of The Acoustical Society Of America, volumen 104, número 5, páginas 3048 a 3058, noviembre de 1998).

20 **[0059]** Dado que el sistema auditivo humano actúa como un analizador de frecuencias con selectividad limitada de frecuencias, se puede incorporar además esta selectividad en frecuencias. Se asume que los filtros auditivos se comportan como filtros de paso banda solapados. En la siguiente explicación a modo de ejemplo, se utiliza un enfoque de banda crítica para aproximar por filtros rectangulares estos pasos banda solapados. El ancho de banda rectangular equivalente (ERB, equivalent rectangular bandwidth) se puede calcular como una función de la frecuencia central (Brian R. Glasberg y Brian C. J. Moore, "Derivation of auditory filter shapes from notched-noise data" (deducción de formas de filtros auditivos a partir de datos de ruido perforado), Hearing Research, volumen 47, páginas 103 a 108, 1990). Considerando que el proceso binaural sigue al filtrado auditivo, ρ se tiene que calcular para canales de frecuencia independientes, produciendo las siguientes señales de presión dependientes de la frecuencia.

35

$$p_{\hat{L}}(n, \omega) = \frac{1}{b(\omega)} \int_{\omega - \frac{b(\omega)}{2}}^{\omega + \frac{b(\omega)}{2}} p_L(n, \omega) d\omega \quad (7)$$

$$p_{\hat{R}}(n, \omega) = \frac{1}{b(\omega)} \int_{\omega - \frac{b(\omega)}{2}}^{\omega + \frac{b(\omega)}{2}} p_R(n, \omega) d\omega, \quad (8)$$

40 donde los límites de integración están dados por los límites de la banda crítica, de acuerdo con la frecuencia central real ω . Los factores $1/b(\omega)$ pueden o no ser utilizados en las ecuaciones (7) y (8).

[0060] Si una de las mediciones de presión del sonido está adelantada o retardada mediante una diferencia de tiempo independiente de la frecuencia, se puede evaluar la coherencia de las señales. El sistema auditivo humano puede hacer uso de dicha propiedad de alineamiento temporal. Habitualmente, la coherencia interaural se calcula dentro de ± 1 ms. Dependiendo de la potencia disponible de procesamiento, los cálculos se pueden implementar utilizando solamente el valor de retardo cero (para baja complejidad), o la coherencia con un adelanto y un retardo temporales (si es posible complejidad elevada). En lo que sigue, no se distinguen ambos casos.

50 **[0061]** El comportamiento ideal se consigue considerando un campo de sonido difuso ideal, que se puede idealizar como un campo de ondas que se compone de ondas planas no correlacionadas, todas con la misma fuerza, que se

propagan en todas las direcciones (es decir, una superposición de un número infinito de ondas planas que se propagan, con relaciones de fase aleatorias y direcciones de propagación distribuidas uniformemente). Una señal radiada mediante un altavoz se puede considerar como una onda plana para un oyente lo suficientemente alejado. Esta hipótesis de onda plana es común en reproducción estereofónica sobre altavoces. Por lo tanto, un campo de sonido sintético reproducido mediante altavoces consiste en ondas planas que contribuyen desde un número limitado de direcciones.

[0062] Dada una señal de entrada con N canales, producida para la reproducción sobre una configuración con posiciones $[l_1, l_2, l_3, \dots, l_N]$ de los altavoces. (En el caso de una configuración de reproducción solamente horizontal, l_i indica el ángulo de acimut. En el caso general, $l_i = (\text{acimut}, \text{elevación})$ indica la posición del altavoz con respecto a la cabeza del oyente. Si la configuración presente en la sala de escucha difiere de la configuración de referencia, l_i puede representar alternativamente las posiciones de los altavoces de la configuración real de reproducción). Con esta información, se puede calcular una curva de referencia de coherencia interaural ρ_{ref} para una simulación de campo difuso, para esta configuración, suponiendo que se alimentan señales independientes a cada altavoz. La potencia de la señal con la que contribuye cada canal de entrada en cada tesela de tiempo-frecuencia se puede incluir en el cálculo de la curva de referencia. En la implementación a modo de ejemplo, ρ_{ref} se utiliza como c_{ref} .

[0063] En las figuras 9a a 9e se muestran diferentes curvas de referencia como ejemplos para curvas de referencia dependientes de la frecuencia o curvas de correlación, para un número diferente de fuentes de sonido en diferentes posiciones de las fuentes de sonido y diferentes orientaciones de la cabeza, tal como se indica en las figuras.

[0064] A continuación, se comenta en mayor detalle el cálculo de los resultados del análisis comentado en el contexto de la figura 8 en base a las curvas de referencia.

[0065] El objetivo es obtener una ponderación igual a 1, si la correlación de los canales de mezcla hacia abajo es igual a la correlación de referencia calculada, suponiendo que se están reproduciendo señales independientes desde todos los altavoces. Si la correlación de la mezcla hacia abajo es igual a +1 ó -1, la ponderación obtenida debería ser 0, lo que indica que no hay componentes independientes presentes. Entre dichos casos extremos, la ponderación debería representar una transición razonable entre la indicación como independiente ($W=1$) o completamente dependiente ($W=0$).

[0066] Dada la curva de correlación de referencia $c_{ref}(\omega)$ y la estimación de la correlación/coherencia de la señal de entrada real reproducida sobre la configuración de reproducción real ($c_{sig}(\omega)$) (c_{sig} es la coherencia de respuesta de correlación de la mezcla hacia abajo), se puede calcular la desviación de $c_{sig}(\omega)$ con respecto a $c_{ref}(\omega)$. La desviación (incluyendo posiblemente un umbral superior y uno inferior) se mapea al intervalo $[0;1]$ para obtener una ponderación ($W(m, i)$) que se aplica a todos los canales de entrada para separar los componentes independientes.

[0067] El ejemplo siguiente muestra un posible mapeo cuando los umbrales se corresponden con la curva de referencia:

[0068] La magnitud de la desviación (indicada como Δ) de la curva real c_{sig} respecto de la c_{ref} de referencia está dada por

$$\Delta(\omega) = |c_{sig}(\omega) - c_{ref}(\omega)| \quad (9)$$

[0069] Dado que la correlación/coherencia está limitada entre $[-1;+1]$, la desviación máximamente posible hacia +1 ó -1 para cada frecuencia está dada por

$$\bar{\Delta}_+(\omega) = 1 - c_{ref}(\omega) \quad (10)$$

$$\bar{\Delta}_-(\omega) = c_{ref}(\omega) + 1 \quad (11)$$

[0070] Por lo tanto, la ponderación para cada frecuencia se obtiene de

$$W(\omega) = \begin{cases} 1 - \frac{\Delta(\omega)}{\bar{\Delta}_+(\omega)} & c_{sig}(\omega) \geq c_{ref}(\omega) \\ 1 - \frac{\Delta(\omega)}{\bar{\Delta}_-(\omega)} & c_{sig}(\omega) < c_{ref}(\omega) \end{cases} \quad (13)$$

[0071] Considerando la dependencia temporal y la resolución limitada de frecuencia de la descomposición de frecuencias, los valores de ponderación se obtienen como sigue (aquí, se proporciona el caso general de una curva de referencia que puede cambiar en el tiempo. Es posible asimismo una curva de referencia independiente del tiempo (es decir, $c_{ref}(i)$):

5

$$W(m, i) = \begin{cases} 1 - \frac{\Delta(m, i)}{\Delta_+(m, i)} & c_{sig}(m, i) \geq c_{ref}(m, i), \\ 1 - \frac{\Delta(m, i)}{\Delta_-(m, i)} & c_{sig}(m, i) < c_{ref}(m, i) \end{cases} \quad (14)$$

[0072] Dicho proceso se puede llevar a cabo en una descomposición de frecuencias, con coeficientes de frecuencia agrupados para sub-bandas motivadas perceptualmente, por razones de complejidad computacional y para obtener filtros con respuestas de impulso más cortas. Además, se podrían aplicar filtros de suavizado y también funciones de compresión (es decir, perturbando la ponderación de manera deseada, introduciendo adicionalmente valores de ponderación mínimo y/o máximo).

[0073] La figura 5 muestra otra implementación de la presente invención, en la que el mezclador hacia abajo está implementado utilizando HRTF y filtros auditivos, tal como se muestra. Además, la figura 5 muestra adicionalmente que los resultados del análisis entregados por el analizador 16 son factores de ponderación para cada sector de tiempo/frecuencia, y el procesador de señal 20 se muestra como un extractor para extraer componentes independientes. A continuación, la salida del procesador 20 son, de nuevo, N canales, pero en este caso cada canal incluye solamente los componentes independientes y ya no incluye componentes dependientes. En esta implementación, el analizador calcularía las ponderaciones de tal modo que, en la primera implementación de la figura 8, un componente independiente recibiría un valor de ponderación de 1 y un componente dependiente de recibiría un valor de ponderación de 0. A continuación, las teselas de tiempo/frecuencia en los N canales originales procesados mediante el procesador 20 que tienen componentes dependientes se configurarían a 0.

[0074] En la otra alternativa en que existen valores de ponderación entre 0 y 1 en la figura 8, el analizador calcularía la ponderación de tal modo que una tesela de tiempo/frecuencia que esté a una pequeña distancia a la curva de referencia recibiría un valor elevado (más próximo a 1), y una tesela de tiempo/frecuencia que esté a una gran distancia a la curva de referencia recibiría un factor de ponderación pequeño (más próximo a 0). En la ponderación mostrada a continuación, por ejemplo, en 20 de la figura 3, los componentes independientes estarían, entonces, amplificados mientras que los componentes dependientes estarían atenuados.

[0075] Sin embargo, cuando el procesador de señal 20 estuviera implementado para no extraer los componentes independientes, para sí extraer los componentes dependientes, entonces las ponderaciones se asignarían a la inversa de tal modo que, cuando la ponderación se realiza en los multiplicadores 20 mostrados en la figura 3, los componentes independientes se atenúan y los componentes dependientes se amplifican. De este modo, cada procesador de señal se puede aplicar para la extracción de los componentes de señal, dado que la determinación de los componentes de señal extraídos realmente está determinada mediante la asignación real de valores de ponderación.

[0076] La figura 6 muestra otra implementación del concepto inventivo, pero en este caso con una implementación diferente del procesador 20. En la realización de la figura 6, el procesador 20 está implementado para extraer "per se" partes difusas independientes, partes directas independientes y partes directas/componentes directos.

[0077] Para obtener, a partir de los componentes independientes separados (Y_1, \dots, Y_N), las partes que contribuyen a la percepción del campo de sonido envolvente/ambiente, es necesario considerar condicionantes adicionales. Uno de dichos condicionantes puede ser la hipótesis de que el sonido ambiente envolvente es igualmente fuerte desde cada dirección. De este modo, por ejemplo, la energía mínima de cada tesela de tiempo-frecuencia en cada canal de las señales de sonido independientes se puede extraer para obtener una señal ambiente envolvente (que se puede procesar adicionalmente para obtener un mayor número de canales ambiente). Ejemplo:

50

$$\tilde{Y}_j(m, i) = g_j(m, i) \cdot Y_j(m, i), \text{ with } g_j(m, i) = \sqrt{\frac{\min_{1 \leq k \leq N} \{P_k(m, i)\}}{P_j(m, i)}}, \quad (15)$$

donde P denota una estimación de potencia de tiempo reducido. (Este ejemplo muestra el caso más simple. Un caso excepcional obvio, donde esto no es aplicable, es cuando uno de los canales incluye pausas de señal durante las cuales la potencia en este canal sería baja o cero.)

[0078] En algunos casos, resulta ventajoso extraer las partes de energía iguales de todos los canales de entrada y

calcular la ponderación utilizando solamente estos espectros extraídos.

$$\tilde{X}_j(m,i) = g_j(m,i) \cdot X_j(m,i), \text{ with } g_j(m,i) = \sqrt{\frac{\min_{1 \leq k \leq N} \{P_{X_k}(m,i)\}}{P_{X_j}(m,i)}}, \quad (16)$$

5 **[0079]** Las partes dependientes extraídas (se pueden obtener, por ejemplo, como $Y_{\text{dependiente}} = Y_j(m,i) - X_j(m,i)$) pueden ser utilizadas para detectar dependencias de los canales y estimar así los indicios direccionales inherentes en la señal de entrada, permitiendo otros procesos tales como, por ejemplo, replanificación.

10 **[0080]** La figura 7 representa una variación del concepto general. La señal de entrada de N canales se alimenta a un generador de señal de análisis (ASG, analysis signal generator). La generación de la señal de análisis de M canales puede incluir, por ejemplo, un modelo de propagación desde los canales/altavoces hasta los oídos, u otros métodos indicados como mezcla hacia abajo a lo largo de este documento. La indicación de los componentes distintos está basada en la señal de análisis. Las máscaras que indican los diferentes componentes se aplican a las señales de entrada (A extracción/D extracción (20a, 20b)). Las señales de entrada ponderadas se pueden procesar
15 adicionalmente (A post / D post (70a, 70b)) para proporcionar señales de salida con carácter específico, donde, en este ejemplo, los indicadores "A" y "D" se han elegido para indicar que los componentes a extraer pueden ser "Ambiente" y "sonido Directo".

20 **[0081]** A continuación se describe la figura 10. Un campo de sonido estacionario se denomina difuso si la distribución direccional de la energía del sonido no depende de la dirección. La distribución direccional de la energía se puede evaluar mediante una medición en todas las direcciones utilizando un micrófono muy direccional. En la acústica de salas, el campo de sonido reverberante en un recinto se modeliza a menudo como un campo difuso. Un campo de sonido difuso se puede idealizar como un campo de ondas que se compone de ondas planas no correlacionadas, todas con la misma fuerza, propagándose en todas las direcciones. Dicho campo de sonido es
25 isótropo y homogéneo.

[0082] Si la uniformidad de la distribución de energía es de interés particular, el coeficiente de correlación punto a punto

$$r = \frac{\langle p_1(t) \cdot p_2(t) \rangle}{[\langle p_1^2(t) \rangle \cdot \langle p_2^2(t) \rangle]^{1/2}}$$

30 de las presiones del sonido en régimen estacionario $p_1(t)$ y $p_2(t)$ en dos puntos separados espacialmente se puede utilizar para estimar la difusión física de un campo de sonido. Para la hipótesis de campos de sonido difusos ideales tridimensionales y bidimensionales en régimen estacionario, inducidos por una fuente sinusoidal, se pueden obtener las relaciones siguientes:

35

$$r_{3D} = \frac{\sin(kd)}{kd},$$

y

$$r_{2D} = J_0(kd),$$

40 donde $k = \frac{2\pi}{\lambda}$ (con λ = longitud de onda) es el número de onda, y d la distancia entre los puntos medidos. Dadas estas relaciones, la difusión de un campo de sonido se puede evaluar comparando datos de mediciones con las curvas de referencia. Dado que las relaciones ideales son condiciones solamente necesarias, pero no suficientes, se puede considerar una serie de mediciones con diferentes orientaciones del eje que conecta los micrófonos.

45 **[0083]** Considerando un oyente en un campo de sonido, las mediciones de la presión de sonido están proporcionadas por las señales de entrada al oído $p_l(t)$ y $p_r(t)$. Por lo tanto, la distancia asumida d entre los puntos de

medición es fija y r pasa a ser una función solamente de la frecuencia con $f = \frac{kc}{2\pi}$, donde c es la velocidad del sonido en el aire. Estas señales de entrada del oído difieren de las señales en campo libre consideradas

anteriormente, debido a la influencia de los efectos provocados por los pabellones auditivos, la cabeza y el torso del oyente. Dichos efectos, sustanciales para la audición espacial, están descritos mediante funciones de transferencia relativas a la cabeza (HRTFs, head-related transfer functions). Se pueden utilizar datos de HRTF medidos para incorporar estos efectos. Se utiliza un modelo analítico para simular una aproximación de las HRTF. La cabeza se modeliza como una esfera rígida con un radio de 8,75 cm y las posiciones de los oídos en un acimut de $\pm 100^\circ$ y una elevación de 0° . Dado el comportamiento teórico de r en un campo de sonido difuso ideal y la influencia de las HRTF, es posible determinar una curva de referencia de correlación cruzada interaural dependiente de la frecuencia para campos de sonido difusos.

10 **[0084]** La estimación del grado de difusión está basada en la comparación de indicios simulados con indicios de referencia del campo difuso asumido. Esta comparación está sujeta a las limitaciones de la audición humana. En el sistema auditivo, el proceso binaural sigue la periferia auditiva consistente en el oído externo, el oído medio y el oído interno. No se consideran los efectos del oído externo que no están aproximados por el modelo esférico (por ejemplo, forma de los pabellones auditivos, conducto del oído) y los efectos del oído medio. La selectividad espectral del oído interior se modeliza como un banco de filtros de paso banda en solapamiento (denominados filtros auditivos en la figura 10). Se utiliza un enfoque de banda crítica para aproximar estos pasos banda en solapamiento mediante filtros rectangulares. Se calcula la anchura rectangular equivalente (ERB, equivalent rectangular bandwidth) como una función de la frecuencia central con arreglo a,

$$b(f_c) = 24.7 \cdot (0.00437 \cdot f_c + 1)$$

20

[0085] Se asume que el sistema auditivo humano es capaz de realizar un alineamiento temporal para detectar componentes de señal coherentes, y que se utiliza análisis de correlación cruzada para la estimación del tiempo de alineamiento τ (correspondiente a ITD) en presencia de sonidos complejos. Hasta aproximadamente 1 a 1,5 kHz, los desplazamientos de tiempo de la señal portadora se evalúan utilizando correlación cruzada de la forma de onda, mientras que a frecuencias superiores la correlación cruzada de la envolvente se convierte en el indicio relevante. En lo que sigue, no se realiza esta distinción. La estimación de la coherencia interaural (IC, interaural coherence) se modeliza como el valor absoluto máximo de la función de correlación cruzada interaural normalizada.

25

$$IC = \max_{\tau} \left| \frac{\langle p_L(t) \cdot p_R(t + \tau) \rangle}{\left[\langle p_L^2(t) \rangle \cdot \langle p_R^2(t) \rangle \right]^{\frac{1}{2}}} \right|$$

30

[0086] Algunos modelos de percepción binaural consideran un análisis de correlación cruzada interaural dinámica. Dado que aquí se consideran señales estacionarias, no se tiene en cuenta la dependencia en el tiempo. Para modelizar la influencia del procesamiento de la banda crítica, se calcula la función de correlación cruzada normalizada dependiente de la frecuencia, como

35

$$IC(f_c) = \frac{\langle A \rangle}{\left[\langle B \rangle \cdot \langle C \rangle \right]^{\frac{1}{2}}}$$

donde A es la función de correlación cruzada por cada banda crítica, y B y C son las funciones de autocorrelación por cada banda crítica. Su relación con el dominio de frecuencia mediante el espectro cruzado de paso banda y los auto-espectros de paso banda se puede formular como sigue:

40

$$A = \max_{\tau} \left| 2 \operatorname{Re} \left(\int_{f_c}^{f_c^*} L^*(f) R(f) e^{j2\pi f(t-\tau)} df \right) \right|,$$

$$B = \left| 2 \left(\int_{f_c}^{f_c^*} L^*(f) L(f) e^{j2\pi f t} df \right) \right|,$$

$$C = \left| 2 \left(\int_{f_c}^{f_c^*} R^*(f) R(f) e^{j2\pi f t} df \right) \right|,$$

$$f^{\pm} = f_c \pm \frac{b(f_c)}{2}$$

donde $L(f)$ y $R(f)$ son las transformadas de Fourier de las señales de entrada del oído, f_c son los límites de integración superior e inferior de la banda crítica de acuerdo con la frecuencia central real, y $*$ indica conjugada compleja.

- 5 **[0087]** Si las señales procedentes de dos o más fuentes en ángulos diferentes se superponen, se provocan indicios ILD y ITD fluctuantes. Dichas variaciones de ILD y ITD en función del tiempo y/o de la frecuencia pueden generar amplitud. Sin embargo, en el promedio a largo plazo, no debe haber ILDs y ITDs en un campo de sonido difuso. Un promedio ITD de cero significa que la correlación entre las señales no se puede aumentar mediante la alineación de tiempos. En principio, las ILDs se pueden evaluar sobre todo el intervalo de frecuencias audibles.
- 10 Debido a que la cabeza no constituye ningún obstáculo a bajas frecuencias, las ILDs son más eficientes a frecuencias medias y altas.

[0088] A continuación se comentan las figuras 11A y 11B para mostrar la implementación alternativa del analizador sin utilizar una curva de referencia, según se comenta en el contexto de la figura 10 o la figura 4.

15 **[0089]** Se aplica una transformada de Fourier de tiempo reducido (STFT) a los canales de audio envolvente de entrada $x_1(n)$ a $x_N(n)$, proporcionando los espectros de tiempo reducido $X_1(m, i)$ a $X_N(m, i)$, respectivamente, donde m es el índice del espectro (tiempo) e i el índice de frecuencia. Se calculan espectros de una mezcla hacia abajo estéreo de la señal de entrada envolvente, designados como $X_1(m, i)$ y $X_2(m, i)$. Para una envolvente 5.1, una mezcla hacia abajo ITU es adecuada como ecuación (1). $X_1(m, i)$ a $X_5(m, i)$ corresponden, por este orden, a los canales izquierdo (L), derecho (R), centro (C), envolvente izquierdo (LS) y envolvente derecho (RS). En adelante, la mayoría de las veces se omiten los índices de tiempo y de frecuencia para abreviar la notación.

20

[0090] En base a las señal estéreo de mezcla hacia abajo, se calculan filtros W_D y W_A para obtener las estimaciones de señal envolvente de sonido directo y ambiente en las ecuaciones (2) y (3).

25

[0091] Con la hipótesis de que la señal de sonido ambiente no está correlacionada entre todos los canales de entrada, se escogen coeficientes de mezcla hacia abajo de tal modo que esta hipótesis se mantiene asimismo para los canales de mezcla hacia abajo. Por lo tanto, se puede formular el modelo de señal de mezcla hacia abajo en la ecuación 4.

30

[0092] D_1 y D_2 representan los espectros STFT de sonido directo correlacionado, y A_1 y A_2 representan sonido ambiente no correlacionado. Se asume además que los sonidos directo y ambiente en cada canal están no correlacionados mutuamente.

35

[0093] La estimación del sonido directo, en el sentido de media cuadrática mínima, se consigue aplicando un filtro de Wiener a la señal envolvente original para suprimir el ambiente. Para tener un solo filtro que se pueda aplicar a todos los canales de entrada, se estiman los componentes directos en la mezcla hacia abajo utilizando el mismo filtro para el canal izquierdo y el derecho que en la ecuación (5).

40

[0094] La función conjunta de error cuadrático medio para la estimación está dada por la ecuación (6).

[0095] $E\{\cdot\}$ es el operador expectativa, y P_D y P_A son las sumas de las estimaciones de potencia a corto plazo de los componentes directo y ambiente (ecuación 7).

45

[0096] La función de error (6) se minimiza haciendo cero su derivada. El filtro resultante para la estimación del sonido directo está en la ecuación 8.

[0097] De manera similar, se puede obtener el filtro de estimación para sonido ambiente, tal como en la ecuación 9.

50

[0098] En lo que sigue, se obtienen estimaciones para P_D y P_A , necesarias para calcular W_D y W_A . La correlación cruzada de la mezcla hacia abajo viene dada por la ecuación 10, donde, dado el modelo de señal de mezcla hacia abajo 4, se hace referencia a (11).

55

[0099] Asumiendo además que los componentes ambiente en la mezcla hacia abajo tienen la misma potencia en los canales de mezcla hacia abajo izquierdo y derecho, se puede escribir la ecuación 12.

[0100] Sustituyendo la ecuación 12 en la última línea de la ecuación 10, y considerando la ecuación 13, se obtienen las ecuaciones (14) y (15).

60

[0101] Tal como se ha comentado en el contexto de la figura 4, la generación de las curvas de referencia para una correlación mínima se puede concebir situando dos o más fuentes de sonido diferentes en una configuración de

reproducción, y situando la cabeza del oyente en cierta posición en esta configuración de reproducción. A continuación, se emiten señales completamente independientes mediante los diferentes altavoces. Para una configuración de dos altavoces, los dos canales tendrían que estar completamente no correlacionados, con una correlación igual a 0 en caso de que no hubiera ningún producto de mezcla cruzada. Sin embargo, estos productos de mezcla cruzada ocurren debido al acoplamiento cruzado desde el lado izquierdo al lado derecho del sistema humano de escucha, y se producen asimismo otros acoplamientos cruzados debido a reverberaciones de la sala, etc. Por lo tanto, las curvas de referencia que se muestran en la figura 4, o en las figuras 9a a 9d no están siempre a 0, sino que tienen valores particularmente diferentes de 0 aunque las señales de referencia imaginadas en este escenario sean completamente independientes. Sin embargo, es importante comprender que realmente no se necesitan estas señales. Es suficiente asimismo suponer una independencia total entre las dos o más señales cuando se calcula la curva de referencia. Sin embargo, en este contexto se debe observar que se pueden calcular otras curvas de referencia para otros escenarios, por ejemplo, utilizando o suponiendo señales que no son completamente independientes, sino que tienen una cierta dependencia o grado de dependencia entre sí conocido previamente. Cuando se calcula dicha curva de referencia diferente, la interpretación o la provisión de factores de ponderación sería diferente con respecto a una curva de referencia en la que se asumieran señales completamente independientes.

[0102] Aunque algunos aspectos se han descrito en el contexto de un aparato, resulta evidente que estos aspectos representan asimismo una descripción del método correspondiente, donde un bloque o dispositivo corresponde a una etapa del método o a una característica de una etapa del método. Análogamente, los aspectos descritos en el contexto de una etapa del método representan asimismo una descripción de un bloque o elemento, o característica correspondiente de un aparato correspondiente.

[0103] La señal descompuesta inventiva se puede almacenar en un medio de almacenamiento digital o se puede transmitir sobre un medio de transmisión, tal como un medio de transmisión inalámbrica o un medio de transmisión cableada, tal como internet.

[0104] Dependiendo de ciertos requisitos de implementación, las realizaciones de la invención se pueden implementar en hardware o en software. La implementación se puede llevar a cabo utilizando un medio de almacenamiento digital, por ejemplo un disco flexible, un DVD, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria flash, que tienen señales de control legibles electrónicamente almacenadas en las mismas, que cooperan (o pueden cooperar) con un sistema informático programable de tal modo que se lleve a cabo el método respectivo.

[0105] Algunas realizaciones acordes con la invención comprenden un soporte de datos no transitorio que tiene señales de control legibles electrónicamente, que pueden cooperar con un sistema informático programable, de tal modo que se lleve a cabo uno de los métodos descritos en el presente documento.

[0106] En general, las realizaciones de la presente invención se pueden implementar como un producto de programa informático con un código de programa, pudiendo funcionar el código de programa para llevar a cabo uno de los métodos cuando el producto de programa informático se ejecuta en un ordenador. El código de programa puede estar almacenado, por ejemplo, en un soporte legible por una máquina.

[0107] Otras realizaciones comprenden el programa informático para llevar a cabo uno de los métodos descritos en el presente documento, almacenado en un soporte legible por una máquina.

[0108] En otras palabras, una realización del método inventivo es, por lo tanto, un programa informático que tiene un código de programa para llevar a cabo uno de los métodos descritos en el presente documento, cuando el programa informático se ejecuta en un ordenador.

[0109] Por lo tanto, una realización adicional de los métodos inventivos es un soporte de datos (o un medio de almacenamiento digital, o un medio legible por ordenador) que comprende, grabado en el mismo, el programa informático para llevar a cabo uno de los métodos descritos en el presente documento.

[0110] Por lo tanto, otra realización del método inventivo es un flujo de datos o una secuencia de señales que representa el programa informático para llevar a cabo uno de los métodos descritos en el presente documento. El flujo de datos o la secuencia de señales puede estar configurado, por ejemplo, para ser transferido a través de una conexión de comunicación de datos, por ejemplo a través de internet.

[0111] Una realización adicional comprende un medio de procesamiento, por ejemplo un procesador, o un dispositivo lógico programable, configurado o adaptado para llevar a cabo uno de los métodos descritos en el presente documento.

[0112] Otra realización comprende un ordenador que tiene instalado en el mismo el programa informático para llevar a cabo uno de los métodos descritos en el presente documento.

[0113] En algunas realizaciones, se puede utilizar un dispositivo lógico programable (por ejemplo, una matriz de puertas programable in situ) para llevar a cabo algunas o todas las funcionalidades de los métodos descritos en el presente documento. En algunas realizaciones, una matriz de puertas programable in situ puede cooperar con un microprocesador para llevar a cabo uno de los métodos descritos en el presente documento. Generalmente, los
5 métodos se llevan a cabo preferentemente mediante cualquier aparato de hardware.

[0114] Las realizaciones descritas anteriormente son meramente ilustrativas de la presente invención. Se comprende que serán evidentes para los expertos en la materia modificaciones y variaciones de las disposiciones y los detalles descritos en el presente documento. Por lo tanto, la intención es limitarse solamente mediante el alcance
10 de las siguientes reivindicaciones de patente y no mediante los detalles específicos presentados a modo de descripción y explicación de las realizaciones del presente documento.

REIVINDICACIONES

1. Aparato para descomponer una señal que tiene una serie de canales, que comprende:
 - 5 un analizador (16) para analizar una similitud entre dos canales de una señal de análisis con respecto a la señal que tiene la serie de canales, a efectos de obtener un resultado del análisis (18), en el que el analizador (16) está configurado para utilizar una curva de similitud dependiente de la frecuencia calculada previamente, como curva de referencia, para determinar el resultado del análisis (18), donde la curva de similitud dependiente de la frecuencia calculada previamente ha sido calculada en base a dos señales para obtener un grado cuantitativo de similitud entre
 - 10 las dos señales sobre un intervalo de frecuencias; y

un procesador de señal (20) para procesar la señal de análisis o una señal obtenida a partir de la señal de análisis, o una señal a partir de la cual se ha obtenido la señal de análisis, utilizando el resultado del análisis para obtener una señal descompuesta.
 - 15 2. Aparato acorde con la reivindicación 1, que comprende además una tabla de consulta en la que está almacenada previamente la curva de referencia.
 3. Aparato acorde con las reivindicaciones 1 ó 2, que comprende además un convertidor de tiempo-frecuencia (32) para convertir la señal o la señal de análisis, o la señal a partir de la cual se ha obtenido la señal de análisis, en una secuencia temporal de representaciones de frecuencia, teniendo cada representación de frecuencia una serie de sub-bandas,
 - 20 en el que el analizador (16) está configurado para determinar, para cada sub-banda, un valor de similitud de referencia a partir de la curva de similitud dependiente de la frecuencia, y para utilizar una similitud entre los dos canales de la sub-banda y el valor de similitud de referencia a efectos de determinar el resultado del análisis para dicha sub-banda.
 4. Aparato acorde con una de las reivindicaciones anteriores, en el que el analizador (16) está
 - 30 configurado para calcular el resultado del análisis comparando un valor de similitud obtenido a partir de los dos canales de la señal de análisis con el valor de similitud correspondiente determinado mediante la curva de referencia, y para asignar un valor de ponderación de acuerdo con el resultado de la comparación o para calcular la diferencia entre el valor de similitud obtenido a partir de los dos canales de la señal de análisis y un valor de similitud correspondiente determinado a partir de la curva de referencia.
 - 35 5. Aparato acorde con una de las reivindicaciones anteriores, en el que el analizador (16) está configurado para producir, como el resultado del análisis, factores de ponderación ($W(m, i)$), y
 - 40 en el que el procesador de señal (20) está configurado para aplicar los factores de ponderación a la señal de entrada o a la señal obtenida a partir de la señal de entrada, mediante una ponderación con los factores de ponderación.
 6. Aparato acorde con una de las reivindicaciones anteriores, que comprende además un mezclador hacia abajo (12) para mezclar hacia abajo una señal de entrada para la señal de análisis, teniendo la señal de entrada más canales que la señal de análisis, y
 - 45 en el que el procesador (20) está configurado para procesar la señal de entrada o una señal obtenida a partir de la señal de entrada diferente de la señal de análisis
 7. Aparato acorde con una de las reivindicaciones anteriores, en el que el analizador (16) está
 - 50 configurado para utilizar la curva de referencia calculada previamente que indica una similitud dependiente de la frecuencia entre dos señales, generada mediante señales que tienen un grado de dependencia conocido previamente.
 8. Aparato acorde con cualquiera de las reivindicaciones anteriores, en el que el analizador está
 - 55 configurado para utilizar una curva de similitud dependiente de la frecuencia almacenada previamente que indica una similitud dependiente de la frecuencia entre dos o más señales en la posición de un oyente, suponiendo que las señales tienen una característica de similitud conocida y que las señales pueden ser emitidas por altavoces en posiciones conocidas de los altavoces.
 - 60 9. Aparato acorde con las reivindicaciones 7 ó 8, en el que se conoce una característica de similitud de la señal de referencia.
 10. Aparato acorde con cualquiera de las reivindicaciones 7, 8 ó 9, en el que las señales de referencia están totalmente decorrelacionadas.
 - 65 11. Aparato acorde con cualquiera de las reivindicaciones anteriores, en el que el analizador (16) está

configurado para analizar canales de mezcla hacia abajo en sub-bandas determinadas mediante una resolución de frecuencia del oído humano.

12. Aparato acorde con cualquiera de las reivindicaciones anteriores, en el que el analizador (16) está configurado para analizar la señal mezclada hacia abajo a efectos de generar un resultado del análisis que permite una descomposición en directa y ambiente, y

en el que el procesador de señal (20) está configurado para extraer la parte directa o la parte de ambiente utilizando el resultado del análisis.

10

13. Aparato acorde con una de las reivindicaciones anteriores, en el que el analizador (16) está configurado para utilizar un límite inferior o superior diferente a la curva de referencia, y en el que el analizador está configurado para comparar un resultado de similitud dependiente de la frecuencia de los canales de análisis con el límite inferior o superior, para determinar el resultado del análisis.

15

14. Método de descomposición de una señal que tiene una serie de canales, que comprende:

analizar (16) una similitud entre dos canales de una señal de análisis en relación con la señal que tiene la serie de canales, utilizando una curva de similitud dependiente de la frecuencia calculada previamente, como una curva de referencia para determinar un resultado del análisis (18), en el que

20

la curva de similitud dependiente de la frecuencia calculada previamente ha sido calculada en base a dos señales para obtener un grado cuantitativo de similitud entre las dos señales sobre un intervalo de frecuencias; y

25 procesar (20) la señal de análisis o una señal obtenida a partir de la señal de análisis, o una señal a partir de la cual se ha obtenido la señal de análisis, utilizando el resultado del análisis para obtener una señal descompuesta.

15. Programa informático para realizar el método de la reivindicación 14, en el que el programa informático es ejecutado por un ordenador o un procesador.

30

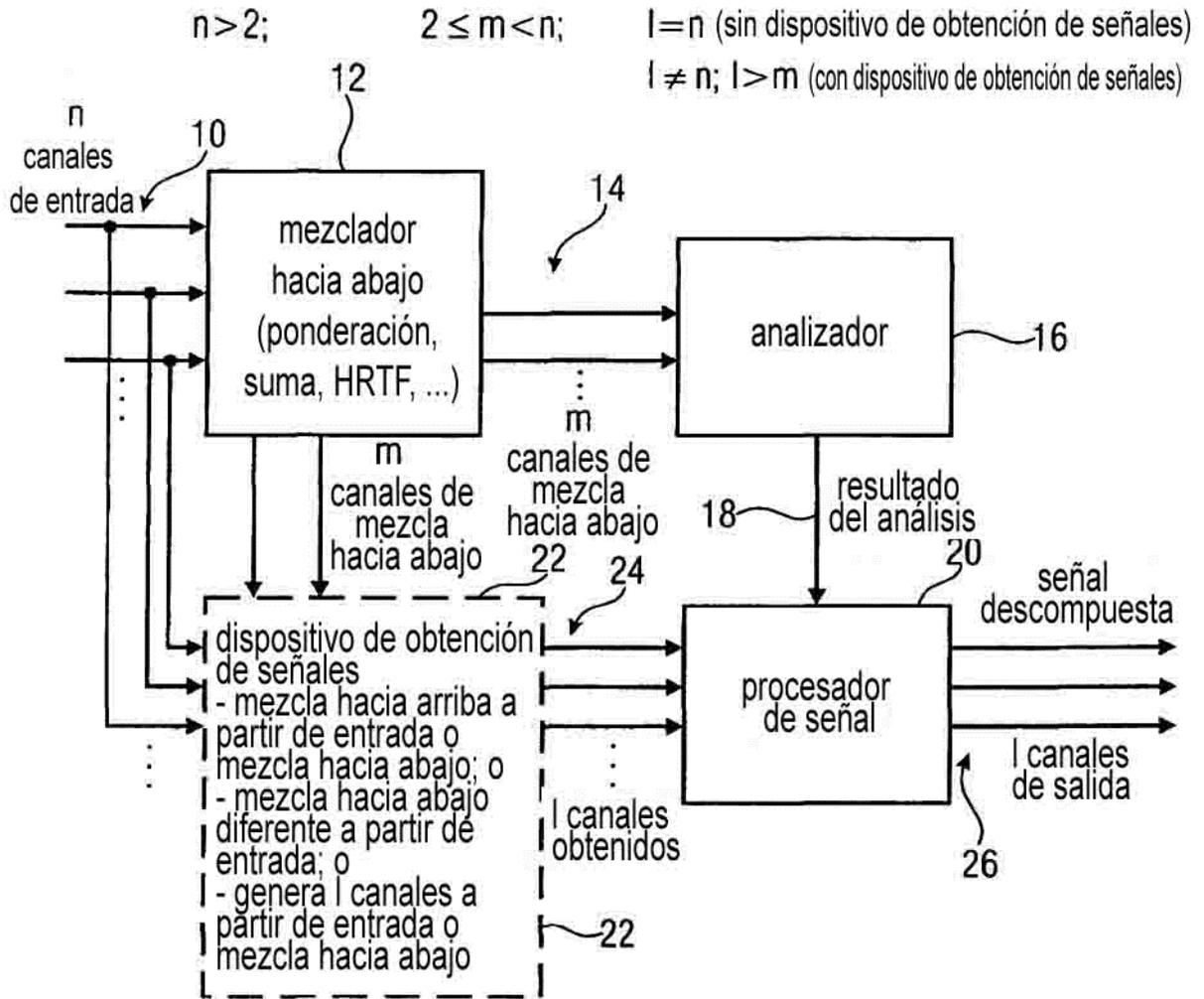


FIG 1

$$i \geq 2; \quad i \neq j; \quad k \geq j$$

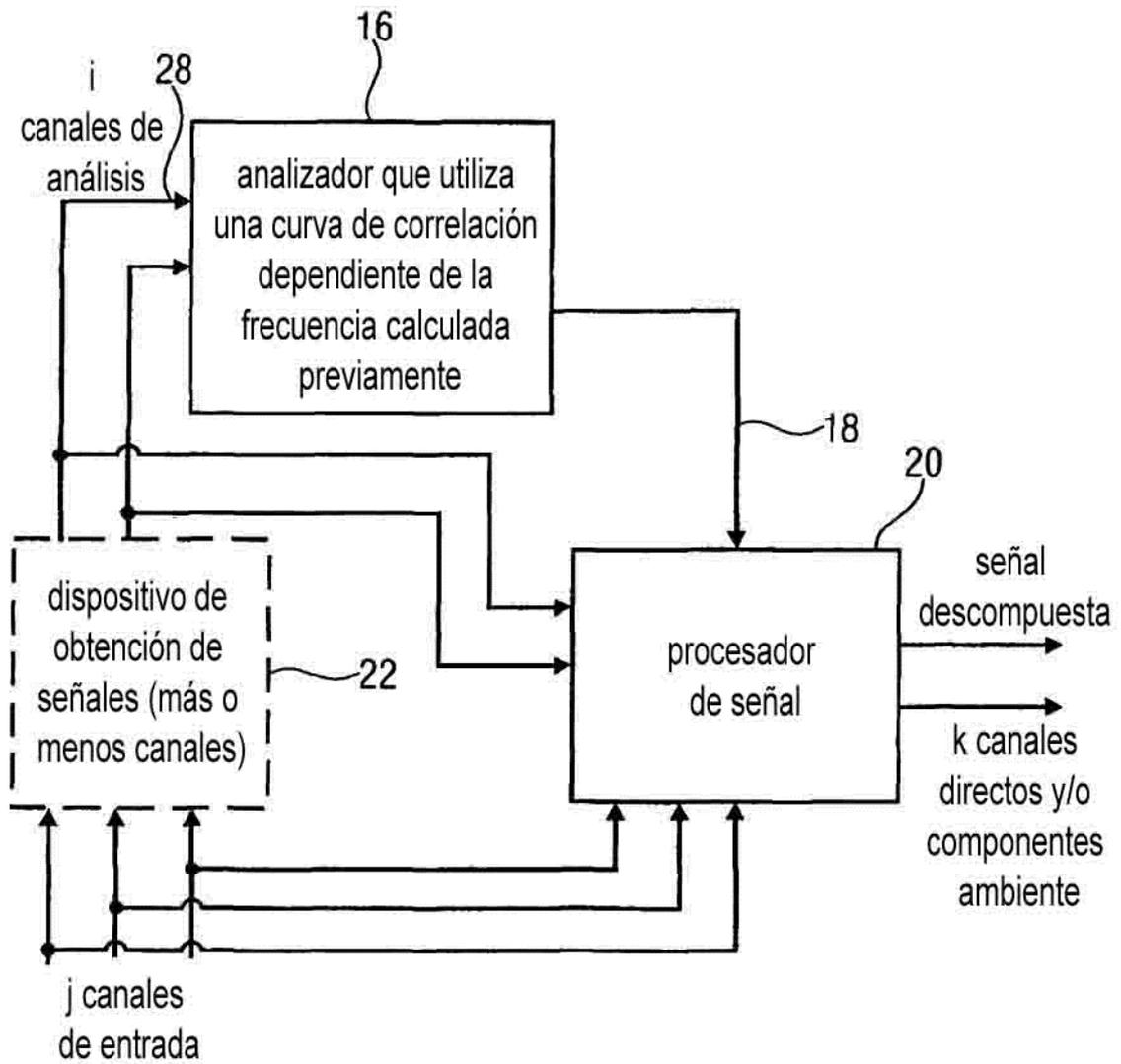


FIG 2

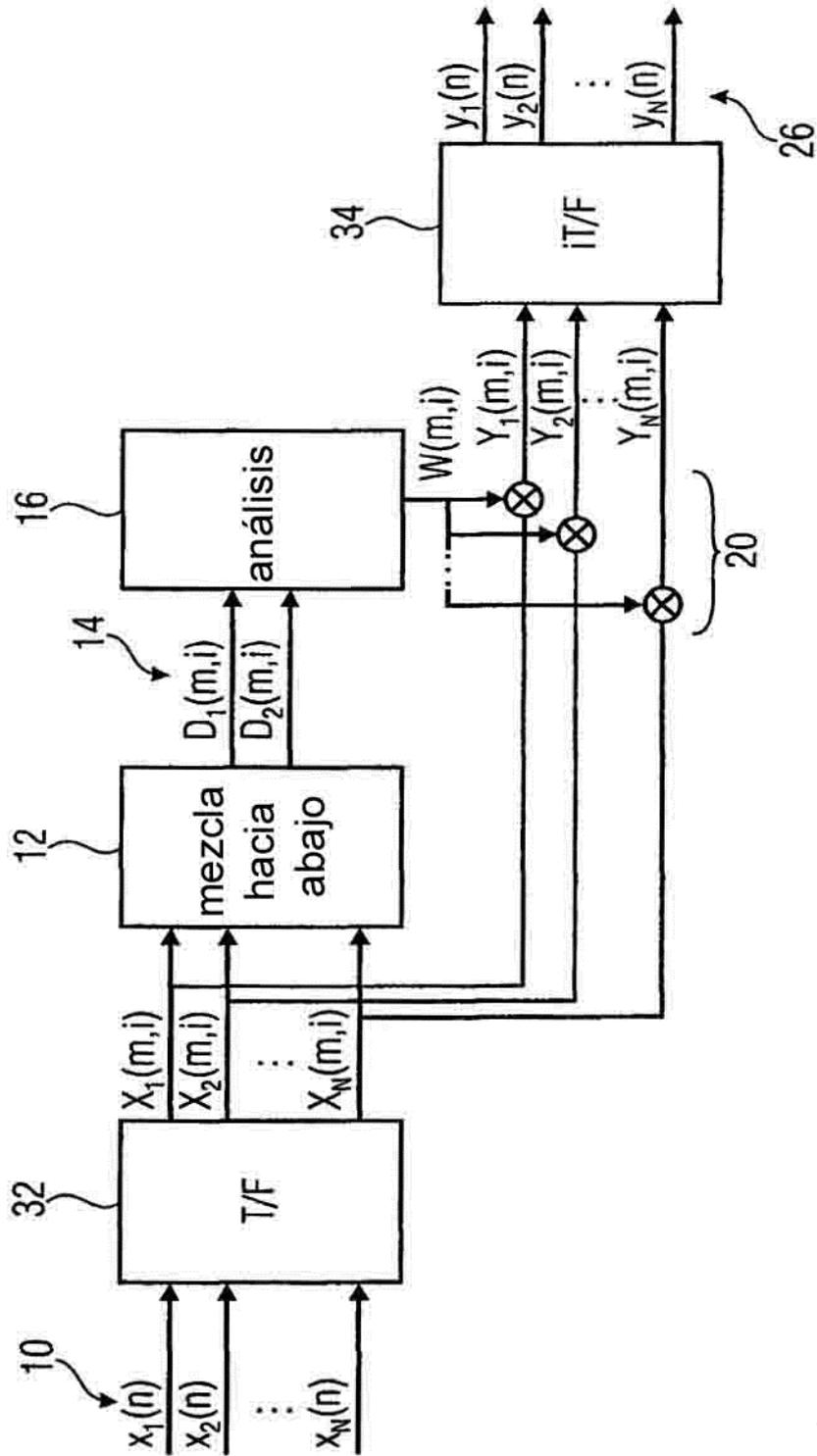


FIG 3

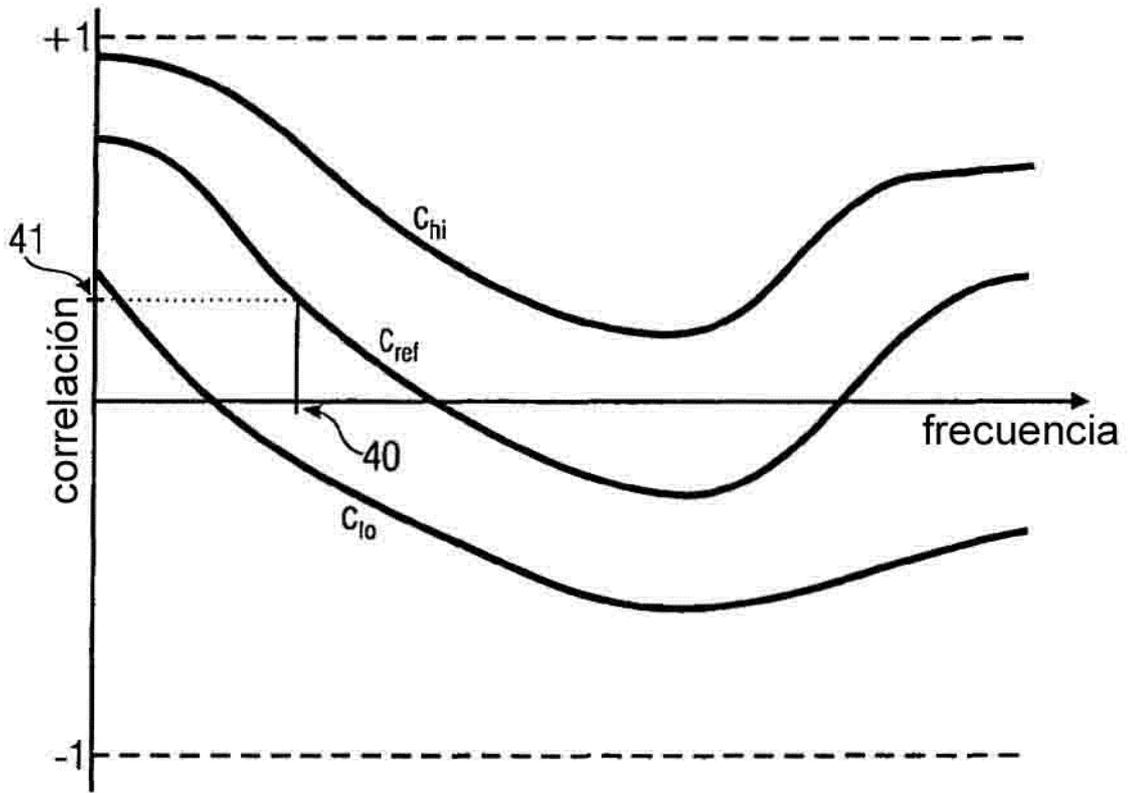


FIG 4

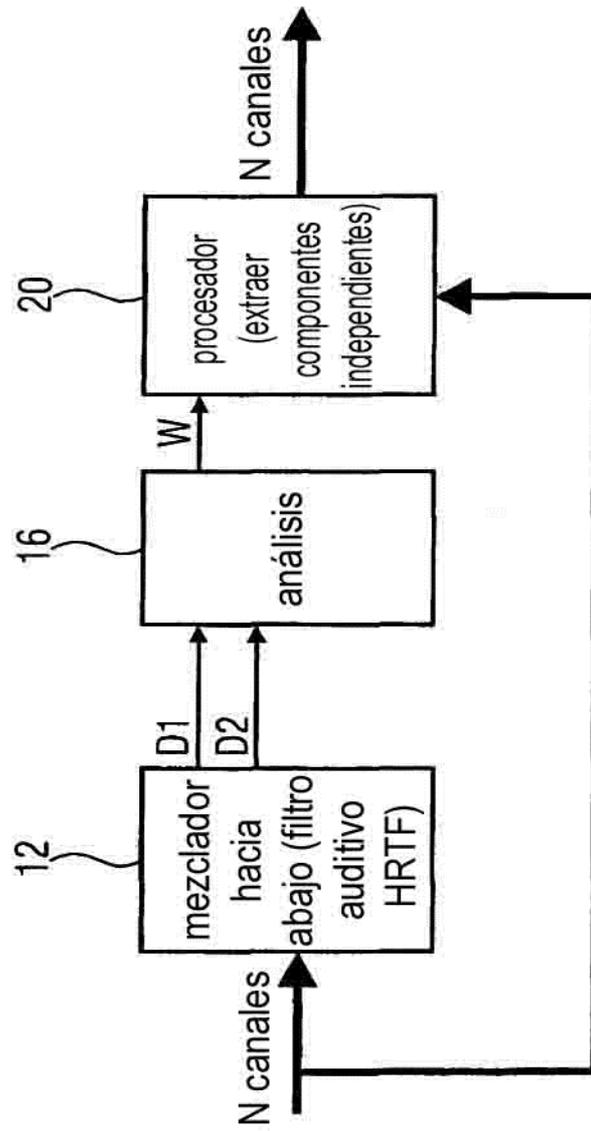


FIG 5

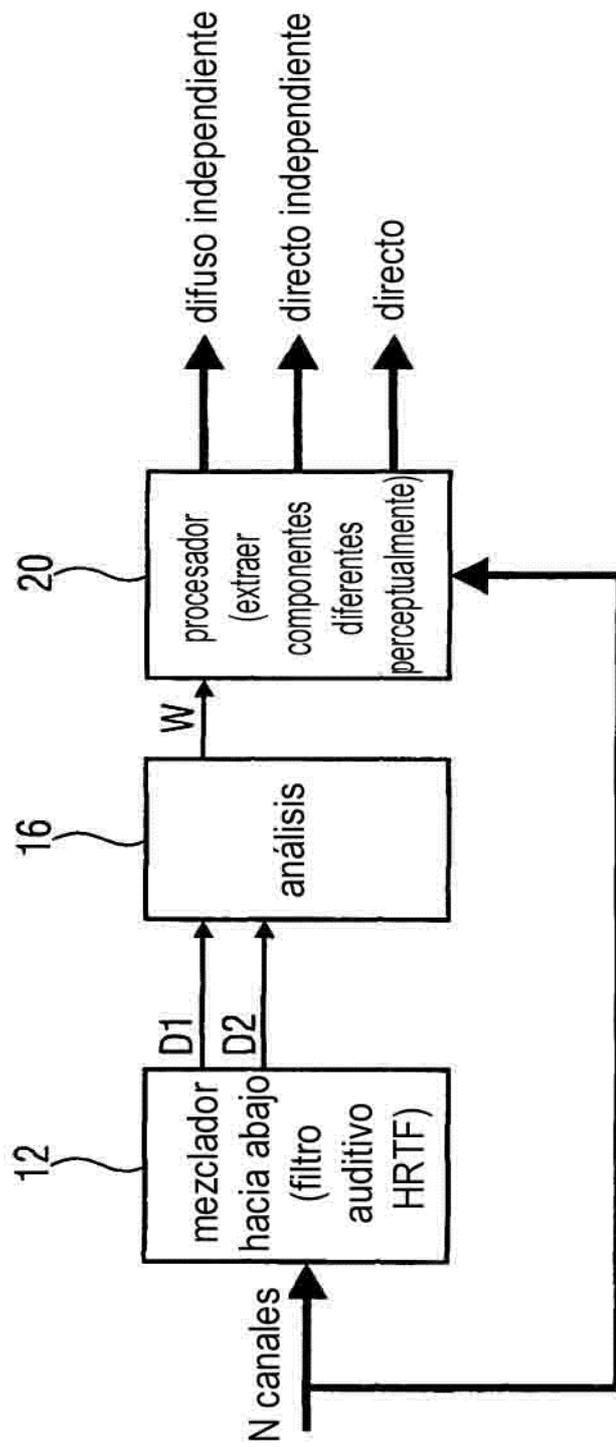


FIG 6

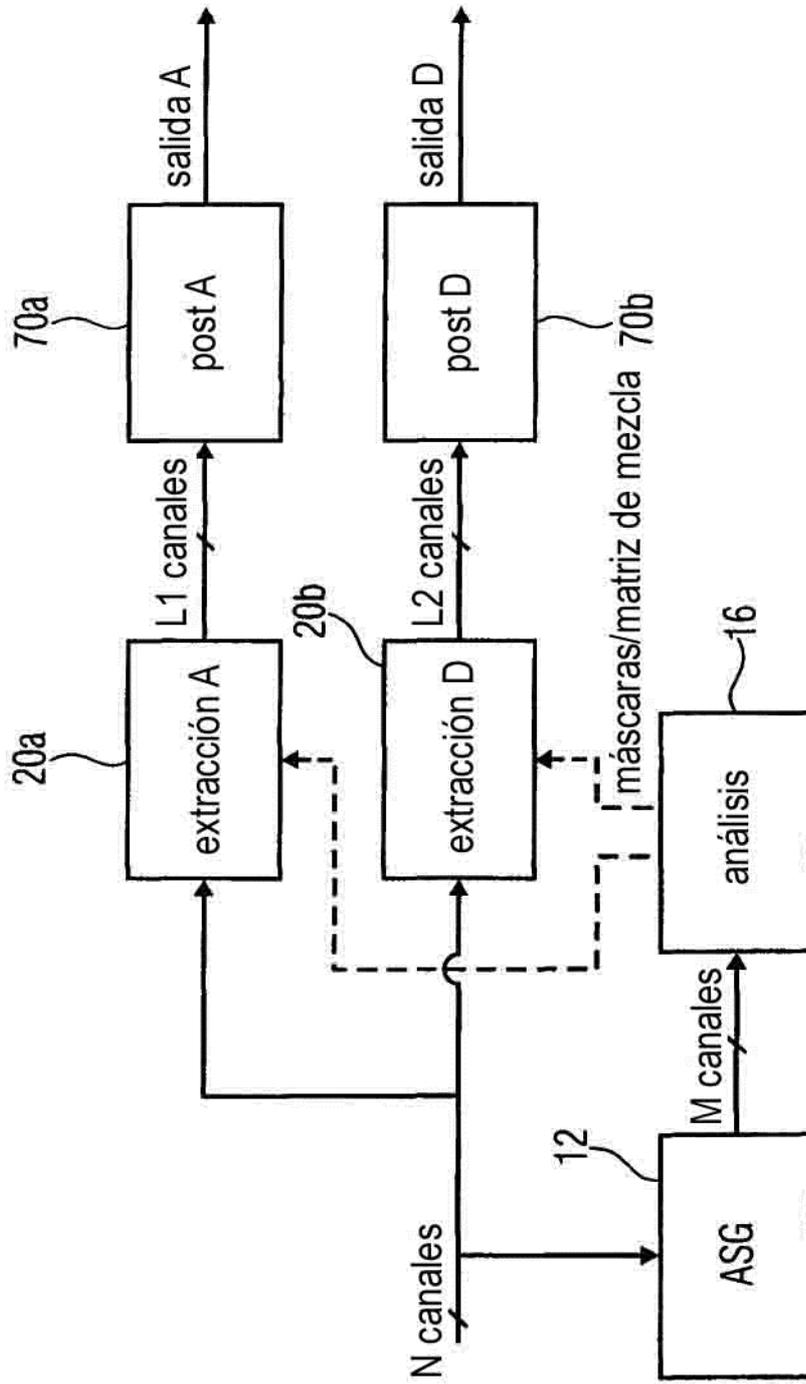


FIG 7

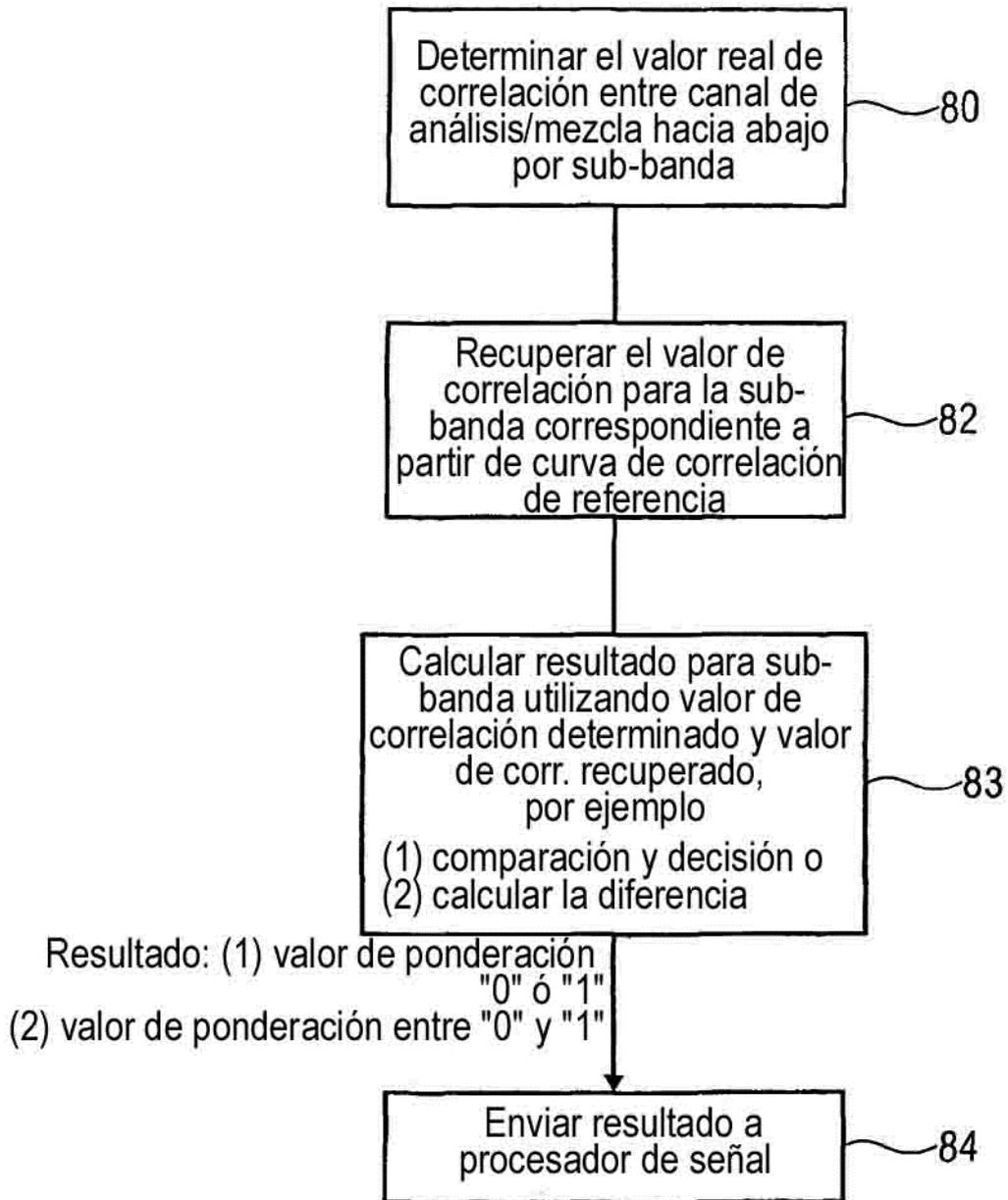
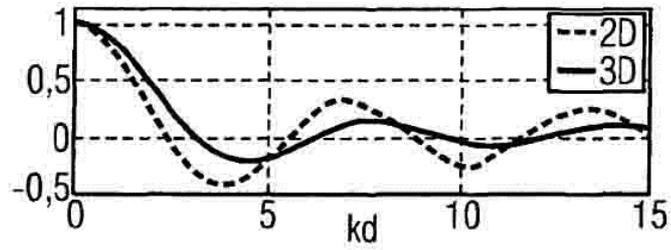
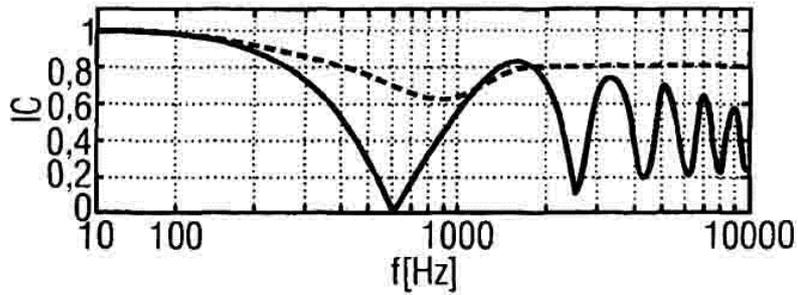


FIG 8



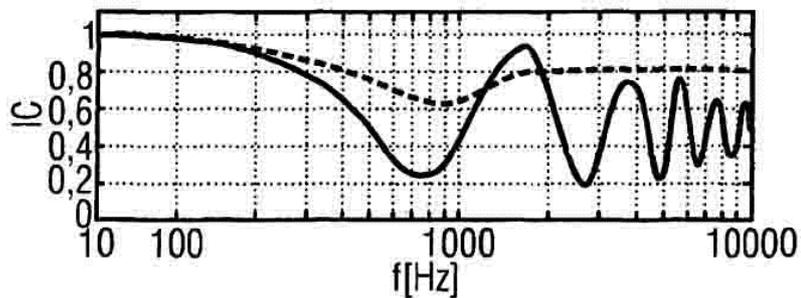
Correlación cruzada (r) como una función de kd en campos de sonido difuso ideal bidimensional y tridimensional

FIG 9A



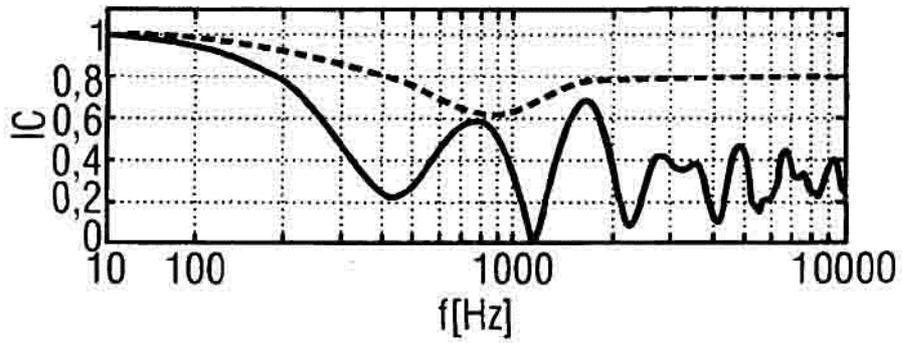
IC (fc) para dos fuentes de sonido en acimut $\pm 30^\circ$ con orientación de la cabeza a 0°

FIG 9B



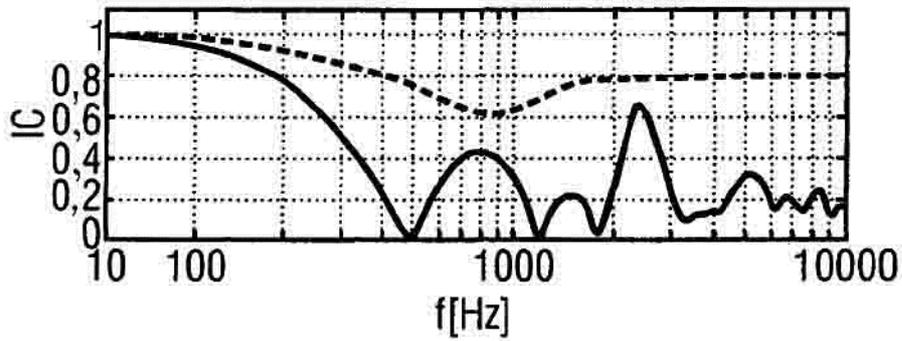
Dos fuentes de sonido en acimut $\pm 30^\circ$, con orientación de la cabeza a 25°

FIG 9C



Cuatro fuentes de sonido en acimut $\{\pm 30^\circ, \pm 110^\circ\}$, con orientación de la cabeza a 25°

FIG 9D



dos fuentes de sonido en acimut $\{0^\circ, \pm 45^\circ, \pm 90^\circ, \pm 135^\circ, \pm 180^\circ\}$, con orientación de la cabeza a 25°

FIG 9E

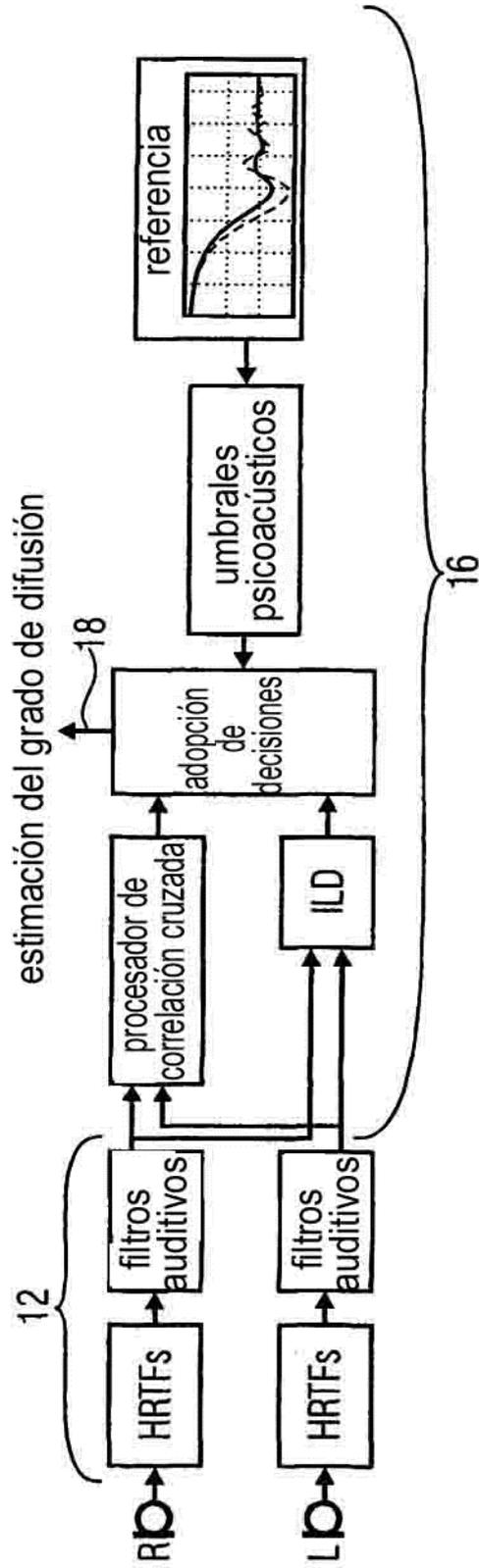


FIG 10

$$\begin{pmatrix} \bar{X}_1(m,i) \\ \bar{X}_2(m,i) \end{pmatrix} = \begin{pmatrix} 1 & 0 & \frac{1}{\sqrt{2}} & 1 & 0 \\ 0 & 1 & \frac{1}{\sqrt{2}} & 0 & 1 \end{pmatrix} \begin{pmatrix} X_1(m,i) \\ \vdots \\ X_5(m,i) \end{pmatrix}, \quad (1)$$

$$\begin{pmatrix} Y_{1D}(m,i) \\ \vdots \\ Y_{ND}(m,i) \end{pmatrix} = W_D(m,i) \begin{pmatrix} X_1(m,i) \\ \vdots \\ X_N(m,i) \end{pmatrix} \quad (2)$$

$$\begin{pmatrix} Y_{1A}(m,i) \\ \vdots \\ Y_{NA}(m,i) \end{pmatrix} = W_A(m,i) \begin{pmatrix} X_1(m,i) \\ \vdots \\ X_N(m,i) \end{pmatrix}, \quad (3)$$

$$\begin{aligned} \bar{X}_1 &= D_1 + A_1 \\ \bar{X}_2 &= D_2 + A_2, \end{aligned} \quad (4)$$

$$\begin{aligned} \hat{D}_1 &= W_D \cdot \bar{X}_1 \\ \hat{D}_2 &= W_D \cdot \bar{X}_2. \end{aligned} \quad (5)$$

$$\begin{aligned} J &= E\{|D_1 - \hat{D}_1|^2\} + E\{|D_2 - \hat{D}_2|^2\} \\ &= (W_d - 1)^2 P_D + W_d^2 P_A, \end{aligned} \quad (6)$$

$$\begin{aligned} P_D &= E\{|D_1|^2\} + E\{|D_2|^2\} \\ P_A &= E\{|A_1|^2\} + E\{|A_2|^2\}. \end{aligned} \quad (7)$$

$$W_D = \frac{P_D}{P_D + P_A}. \quad (8)$$

$$W_A = \frac{P_A}{P_D + P_A} = 1 - W_D. \quad (9)$$

FILTRADO DE WIENER PARA ANÁLISIS

FIG 11A

$$\begin{aligned}
\operatorname{Re}(E\{\check{X}_1\check{X}_2^*\}) &= \operatorname{Re}(E\{(D_1+A_1)(D_2^*+A_2^*)\}) \\
&= \operatorname{Re}(E\{D_1D_2^*\}) \\
&= \sqrt{E\{|D_1|^2\}E\{|D_2|^2\}}.
\end{aligned} \tag{10}$$

$$\begin{aligned}
E\{|D_1|^2\} &= E\{|\bar{X}_1|^2\} - E\{|A_1|^2\} \\
E\{|D_2|^2\} &= E\{|\bar{X}_2|^2\} - E\{|A_2|^2\}.
\end{aligned} \tag{11}$$

$$\begin{aligned}
E\{|D_1|^2\} &= E\{|\bar{X}_1|^2\} - \frac{P_A}{2} \\
E\{|D_2|^2\} &= E\{|\bar{X}_2|^2\} - \frac{P_A}{2}.
\end{aligned} \tag{12}$$

$$P_D + P_A = E\{|\bar{X}_1|^2\} + E\{|\bar{X}_2|^2\}, \tag{13}$$

$$P_D = \sqrt{(E\{|\bar{X}_1|^2\} - E\{|\bar{X}_2|^2\})^2 + 4\operatorname{Re}(E\{\bar{X}_1\bar{X}_2^*\})^2} \tag{14}$$

$$P_A = E\{|\bar{X}_1|^2\} + E\{|\bar{X}_2|^2\} - P_D. \tag{15}$$

FILTRADO DE WIENER PARA ANÁLISIS

FIG 11B